Clp-protease as target for herbicides

The present invention relates to Clp-protease, which, when absent, brings about reduced growth and chlorotic leaves as target for herbicides. For this purpose, novel nucleic acid sequences encompassing SEQ ID NO:3, SEQ ID NO:11 and SEQ ID NO: 17 and functional equivalents of SEQ ID NO:3, SEQ ID NO:11 and SEQ ID NO: 17 are provided. Moreover, the present invention relates to the use of Clp-protease in a method for identifying compounds with herbicidal or growth-regulatory activity, and to the use of the compounds identified by this method as herbicides or growth regulators.

The basic principle of identifying herbicides via the inhibition of a defined target is known (for example US 5,187,071, WO 98/33925, WO 00/77185). In general, there is a great demand for the detection of enzymes which might constitute novel targets for herbicides. The reasons are resistance problems which occur with herbicidal active ingredients which act on known targets, and the ongoing endeavor to identify novel herbicidal active ingredients which are distinguished by as wide as possible a spectrum of action, ecological and toxicological acceptability and/or low application rates.

In practice, the detection of novel targets entails great difficulties since the inhibition of an enzyme which forms part of a metabolic pathway frequently has no further effect on the growth of the plant. This may be attributed to the fact that the plant switches to alternative metabolic pathways whose existence is not known or that the inhibited enzyme is not limiting for the metabolic pathway. Furthermore, plant genomes are distinguished by a high degree of functional redundancy. Functionally equivalent enzymes are found more frequently in gene families in the Arabidopsis thaliana genome than in insects or mammals (Nature, 2000, 408(6814):796-815). This hypothesis is confirmed experimentally by the fact that comprehensive gene knock-out programs by T-DNA or transposon insertion into Arabidopsis yielded fewer manifested phenotypes to date than expected (Curr. Op. Plant Biol. 4, 2001, pp.111-117).

It is an object of the present invention to identify novel targets which are essential for the growth of plants or whose inhibition leads to reduced plant growth, and to provide methods which are suitable for identifying herbicidally active and/or growth-regulatory compounds.

We have found that this object is achieved by the use of nuclear encoded Clp-protease in a method for identifying herbicides.

Further terms used in the description are now defined at this point.

"Affinity tag": this refers to a peptide or polypeptide whose coding nucleic acid sequence can be fused to the nucleic acid sequence according to the invention either

directly or by means of a linker, using customary cloning techniques. The affinity tag serves for the isolation, concentration and/or selective purification of the recombinant target protein by means of affinity chromatography from total cell extracts. The abovementioned linker can advantageously contain a protease cleavage site (for example for thrombin or factor Xa), whereby the affinity tag can be cleaved from the target protein when required. Examples of common affinity tags are the "His tag", for example from Qiagen, Hilden, "Strep tag", the "Myc tag" (Invitrogen, Carlsberg), the tag from New England Biolabs which consists of a chitin-binding domain and an inteine, the maltose-binding protein (pMal) from New England Biolabs, and what is known as the CBD tag from Novagen. In this context, the affinity tag can be attached to the 5' or the 3' end of the coding nucleic acid sequence with the sequence encoding the target protein.

"Activity of nuclear encoded Clp-protease": the term activity describes the ability of an enzyme to convert a substrate into a product. The enzymatic activity can be determined in what is known as an activity assay via the increase in the product, the decrease in the substrate (or starting material) or the decrease in a specific cofactor, or via a combination of at least two of the abovementioned parameters, as a function of a defined period of time. "Activity of nuclear encoded Clp-protease" describes here the ability of an enzyme to catalyze the hydrolysis of peptides of maximal five amino acids in vitro.

"Expression cassette": an expression cassette contains a nucleic acid sequence according to the invention linked operably to at least one genetic control element, such as a promoter, and, advantageously, a further control element, such as a terminator. The nucleic acid sequence of the expression cassette can be for example a genomic or complementary DNA sequence or an RNA sequence, and their semisynthetic or fully synthetic analogs. These sequences can exist in linear or circular form, extrachromosomally or integrated into the genome. The nucleic acid sequences in question can be synthesized or obtained naturally or contain a mixture of synthetic and natural DNA components, or else consist of various heterologous gene segments of various organisms.

Artificial nucleic acid sequences are also suitable in this context as long as they make possible the expression, in a cell or an organism, of a polypeptide with the enzymatic activity of a nuclear encoded Clp Protease, preferably with the biological activity of a a nuclear encoded Clp Protease, which polypeptide is encoded by a nucleic acid sequence according to the invention. For example, synthetic nucleotide sequences can be generated which have been optimized with regard to the codon usage of the organisms to be transformed.

All of the abovementioned nucleotide sequences can be generated from the nucleotide units by chemical synthesis in the manner known per se, for example by fragment con-

densation of individual overlapping complementary nucleotide units of the double helix. Oligonucleotides can be synthesized chemically for example in the manner known per se using the phosphoamidite method (Voet, Voet, 2nd Edition, Wiley Press New York, pp. 896-897). When preparing an expression cassette, various DNA fragments can be manipulated in such a way that a nucleotide sequence with the correct direction of reading and the correct reading frame is obtained. The nucleic acid fragments are linked with each other via general cloning techniques as are described, for example, in T. Maniatis, E.F. Fritsch and J. Sambrook, "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989), in T.J. Silhavy, M.L. Berman and L.W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984) and in Ausubel, F.M. et al., "Current Protocols in Molecular Biology", Greene Publishing Assoc. and Wiley-Interscience (1994).

"Operable linkage" or "functional linkage": an operable, or functional, linkage is understood as meaning the sequential arrangement of regulatory sequences or genetic control elements in such a way that each of the regulatory sequences, or each of the genetic control elements, can fulfill its intended function when the coding sequence is expressed.

"Functional equivalents" describe, in the present context, nucleic acid sequences which hybridize under standard conditions with the nucleic acid sequence SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:3, SEQ ID NO:13, SEQ ID NO:15 or SEQ ID NO:17 or parts of the aforementioned nucleic acid sequences and which are capable of bringing about the expression, in a cell or an organism, of a polypeptide with the activity of Clp protease.

To carry out the hybridization, it is advantageous to use short oligonucleotides with a length of approximately 10-50 bp, preferably 15-40 bp, for example of the conserved or other regions, which can be determined in the manner with which the skilled worker is familiar by comparisons with other related genes. However, longer fragments of the nucleic acids according to the invention with a length of 100-500 bp, or the complete sequences, may also be used for hybridization. Depending on the nucleic acid/oligonucleotide used, the length of the fragment or the complete sequence, or depending on which type of nucleic acid, i.e. DNA or RNA, is being used for the hybridization, these standard conditions vary. Thus, for example, the melting temperatures for DNA:DNA hybrids are approximately 10°C lower than those of DNA:RNA hybrids of the same length.

Standard hybridization conditions are to be understood as meaning, depending on the nucleic acid, for example temperatures of between 42 and 58oC in an aqueous buffer solution with a concentration of between 0.1 and 5 x SSC (1 X SSC = 0.15 M NaCl, 15 mM sodium citrate, pH 7.2) or additionally in the presence of 50% formamide, such as,

for example, 42°C in 5 x SSC, 50% formamide. The hybridization conditions for DNA:DNA hybrids are advantageously 0.1 x SSC and temperatures of between approximately 20°C and 65°C, preferably between approximately 30°C and 45°C. In the case of DNA:RNA hybrids, the hybridization conditions are advantageously 0.1 x SSC and temperatures of between approximately 30 °C and 65 °C, preferably between approximately 45°C and 55 °C. These hybridization temperatures which have been stated are melting temperature values which have been calculated by way of example for a nucleic acid with a length of approx. 100 nucleotides and a G + C content of 50% in the absence of formamide. The experimental conditions for DNA hybridization are described in relevant textbooks of genetics such as, for example, in Sambrook et al., "Molecular Cloning", Cold Spring Harbor Laboratory, 1989, and can be calculated using formulae with which the skilled worker is familiar, for example as a function of the length of nucleic acids, the type of the hybrids or the G + C content. The skilled worker will find further information on hybridization in the following textbooks: Ausubel et al. (eds), 1985, "Current Protocols in Molecular Biology", John Wiley & Sons, New York; Hames and Higgins (eds.), 1985, "Nucleic Acids Hybridization: A Practical Approach", IRL Press at Oxford University Press, Oxford; Brown (ed.), 1991, Essential Molecular Biology: A Practical Approach, IRL Press at Oxford University Press, Oxford.

A functional equivalent of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:3, SEQ ID NO:13, SEQ ID NO:15 or SEQ ID NO:17 can be furthermore defined by the degree of homology or identity with SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:3, SEQ ID NO:13, SEQ ID NO:15 or SEQ ID NO:17, respectively, and can furthermore comprise also natural or artificial mutations of the aforementioned nucleic acid sequences which encode a polypeptide with the activity of a nuclear encoded Clp-protease.

The present invention also encompasses, for example, those nucleotide sequences which are obtained by modification of the SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:3, SEQ ID NO:13, SEQ ID NO:15 or SEQ ID NO:17.

For example, such modifications can be generated by techniques with which the skilled worker is familiar, such as "Site Directed Mutagenesis", "Error Prone PCR", "DNA-shuffling" (Nature 370, 1994, pp.389-391) or "Staggered Extension Process" (Nature Biotechnol. 16, 1998, pp.258-261). The aim of such a modification can be, for example, the insertion of further cleavage sites for restriction enzymes, the removal of DNA in order to truncate the sequence, the substitution of nucleotides to optimize the codons, or the addition of further sequences. Proteins which are encoded via modified nucleic acid sequences must retain the desired function despite a deviating nucleic acid sequence.

The term "functional equivalents" can also relate to the amino acid sequence encoded by the nucleic acid sequence in question. In this case, the term "functional equivalent" describes a protein whose amino acid sequence has a defined percentage of identity or homology with SEQ ID NO:3.

Functional equivalents thus also comprise naturally occurring variants of the herein-described sequences and artificial nucleic acid sequences, for example those which have been obtained by chemical synthesis and which are adapted to the codon usage, and also the amino acid sequences derived from them.

"Genetic control sequence" describes sequences which have an effect on the transcription and, if appropriate, translation of the nucleic acids according to the invention in prokaryotic or eukaryotic organisms. Examples thereof are promoters, terminators or what are known as "enhancer" sequences. In addition to these control sequences, or instead of these sequences, the natural regulation of these sequences may still be present before the actual structural genes and may, if appropriate, have been genetically modified in such a way that the natural regulation has been switched off and the expression of the target gene has been modified, that is to say increased or reduced. The choice of the control sequence depends on the host organism or starting organism. Genetic control sequences furthermore also comprise the 5'-untranslated region, introns or the noncoding 3'-region of genes. Control sequences are furthermore understood as meaning those which make possible homologous recombination or insertion into the genome of a host organism or which permit removal from the genome. Genetic control sequences also comprise further promoters, promoter elements or minimal promoters, and sequences which have an effect on the chromatin structure (for example matrix attachment regions (MARs)), which can modify the expression-governing properties. Thus, genetic control sequences may bring about for example the additional dependence of the tissue-specific expression on certain stress factors. Such elements have been described, for example, for water stress, abscísic acid (Lam E and Chua NH, J Biol Chem 1991; 266(26): 17131 -17135), high- and low-temperature stress (Plant Cell 1994, (6): 251-264) and heat stress (Molecular & General Genetics, 1989, 217(2-3): 246-53).

"Homology" between two nucleic acid sequences or polypeptide sequences is defined by the identity of the nucleic acid sequence/polypeptide sequence over in each case the entire sequence length, which is calculated by alignment with the aid of the program algorithm GAP according to Needleman and Wunsch 1970, J. Mol. Biol. 48; 443-453) setting the following parameters for polypeptides:

Gap Weight: 8 Length Weight: 2
Average Match: 2,912 Average Mismatch:-2,003

and the following parameters for nucleic acids:

Gap Weight: 50 Length Weight: 3
Average Match: 10.000 Average Mismatch: 0.000

In the following text, the term identity is also used synonymously with the term "homology".

"Mutations" of nucleic or amino acid sequences comprise substitutions, additions, deletions, inversions or insertions of one or more nucleotide residues, which may also bring about changes in the corresponding amino acid sequence of the target protein by substitution, insertion or deletion of one or more amino acids, although the functional properties of the target proteins are, overall, essentially retained.

"Natural genetic environment" means the natural chromosomal locus in the organism of origin. In the case of a genomic library, the natural genetic environment of the nucleic acid sequence is preferably retained at least in part. The environment flanks the nucleic acid sequence at least at the 5'- or 3'-side and has a sequence length of at least 50 bp, preferably at least 100 bp, especially preferably at least 500 bp, very especially preferably at least 1000 bp, and most preferably at least 5000 bp.

"Plants" for the purposes of the invention are plant cells, plant tissues, plant organs, or intact plants, such as seeds, tubers, flowers, pollen, fruits, seedlings, roots, leaves, stems or other plant parts. Moreover, the term plants is understood as meaning propagation material such as seeds, fruits, seedlings, slips, tubers, cuttings or root stocks.

"Recombinant DNA" describes a combination of DNA sequences which can be generated by recombinant DNA technology.

"Recombinant DNA technology": generally known techniques for fusing DNA sequences (for example described in Sambrook et al., 1989, Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press).

"Replication origins" ensure the multiplication of the expression cassettes or vectors according to the invention in microorganisms and yeasts, for example the pBR322 ori or the P15A ori in E. coli (Sambrook et al.: "Molecular Cloning. A Laboratory Manual", 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989) and the ARS1 ori in yeast (Nucleic Acids Research, 2000, 28(10): 2060-2068).

"Reporter genes" encode readily quantifiable proteins. The transformation efficacy or the expression site or timing can be assessed by means of these genes via growth

assay, fluorescence assay, chemoluminescence assay, bioluminescence assay or resistance assay or via a photometric measurement (intrinsic color) or enzyme activity. Very especially preferred in this context are reporter proteins (Schenborn E, Groskreutz D. Mol Biotechnol. 1999; 13(1):29-44) such as the "green fluorescent protein" (GFP) (Gerdes HH and Kaether C, FEBS Lett. 1996; 389(1):44-47; Chui WL et al., Curr Biol 1996, 6:325-330; Leffel SM et al., Biotechniques. 23(5):912-8, 1997), chloramphenicol acetyl transferase, a luciferase (Giacomin, Plant Sci 1996, 116:59-72; Scikantha, J Bact 1996, 178:121; Millar et al., Plant Mol Biol Rep 1992 10:324-414), and luciferase genes, in general β-galactosidase or β-glucuronidase (Jefferson et al., EMBO J. 1987, 6, 3901-3907) or the Ura3 gene.

"Selection markers" confer resistance to antibiotics or other toxic compounds: examples which may be mentioned in this context are the neomycin phosphotransferase gene, which confers resistance to the aminoglycoside antibiotics neomycin (G 418), kanamycin, paromycin (Deshayes A et al., EMBO J. 4 (1985) 2731-2737), the sul gene, which encodes a mutated dihydropteroate synthase (Guerineau F et al., Plant Mol Biol. 1990; 15(1):127-136), the hygromycin B phosphotransferase gene (Gen Bank Accession NO: K 01193) and the shble resistance gene, which confers resistance to the bleomycin antibiotics such as zeocin. Further examples of selection marker genes are genes which confer resistance to 2-deoxyglucose-6-phosphate (WO 98/45456) or phosphinothricin and the like, or those which confer a resistance to antimetabolites, for example the dhfr gene (Reiss, Plant Physiol. (Life Sci. Adv.) 13 (1994) 142-149). Examples of other genes which are suitable are trpB or hisD (Hartman SC and Mulligan RC, Proc Natl Acad Sci U S A. 85 (1988) 8047-8051). Another suitable gene is the mannose phosphate isomerase gene (WO 94/20627), the ODC (ornithine decarboxylase) gene (McConlogue, 1987 in: Current Communications in Molecular Biology, Cold Spring Harbor Laboratory, Ed.) or the Aspergillus terreus deaminase (Tamura K et al., Biosci Biotechnol Biochem. 59 (1995) 2336-2338).

"Transformation" describes a process for introducing heterologous DNA into a pro- or eukaryotic cell. The term transformed cell describes not only the product of the transformation process per se, but also all of the transgenic progeny of the transgenic organism generated by the transformation.

"Target/target protein": a polypeptide encoded via the nucleic acid sequence according to the invention (this term is defined herein below), which may take the form of an enzyme in the traditional sense or, for example, of a structural protein, a protein relevant for developmental processes, regulatory protein such as transcription factors, kinases, phosphatases, receptors, channel subunits, transport proteins, regulatory subunits which confer substrate or activity regulation to an enzyme complex. All of the targets or sites of action share the characteristic that their functional presence is essential for survival or normal development and growth.

"Transgenic": referring to a nucleic acid sequence, an expression cassette or a vector comprising a nucleic acid sequence according to the invention or an organism transformed with the abovementioned nucleic acid sequence, expression cassette or vector, the term transgenic describes all those constructs which have been generated by genetic engineering methods in which either the nucleic acid sequence of the target protein or a genetic control sequence linked operably to the nucleic acid sequence of the target protein or a combination of the abovementioned possibilities are not in their natural genetic environment or have been modified by recombinant methods. In this context, the modification can be achieved, for example, by mutating one or more nucleotide residues of the nucleic acid sequence in question.

Intracellular protein degradation and its regulation are important for several processes like recycling of aminoacids, prevention of protein agglomeration and regulation of signaling processes (e.g. signalling of phytohormones). Cytosolic Proteins that are to be degraded are ubiquitinylated at the N-terminus and delivered to the proteasome which is established as a complex of a large number of protein components in eucaroytes. Roughly 12% of the genes in Arabidopsis thaliana are encoding proteins envolved in protein degradation by the ubiquitin pathway.

The stroma of plant chloroplasts contains a unique ubiquitin-independent, ATP-dependent protease consisting of two mayor components, a serine-type protease (ClpP) and an ATPase (ClpC, -D, -X) both of which are encoded by enzyme families in Arabidopsis thaliana (for details on the differing nomenclatures in literature see Adam et al. 2001, Plant Physiology 125, pp.1912-18). Six unique ClpP Isoforms (ClpP1-6) are nuclear encoded in Arabidopsis and at least one ClpP ist encoded in the plastid genome (pClpP) all of which carry the three conserved active site aminoacids characteristical for a catalytic triade of serine proteases. *Some sequences of mRNA for putative ATP-dependent protease proteolytic subunits ClpP are disclosed in Nakabayashi et al. (Plant Cell Physiol 40: 504–514, 1999) and Kotani et al. (DNA Research 4, 291-300, 1997). A subunit of Clp protease, which does not show any own activbity of a protease is disclosed in WO 2003008440 A. Further Clp gene from algae, tobacco or cyanobacterium are depicted in Huang et al. (Mol. Gen. Genet 244, 151-159, 1994), Shikanai et al. (Plant Cell Physiol. 42, 261-273, 2001) and Clarke et al. (Plant Molecular Biology 37, 791-801, 1998) respectivelly.*
Further three nuclear encoded ClpP-Isoforms which miss the conserved amino acid residues of the catalytic triade are found in Arabidopsis (ClpR1, ClpR3, ClpR4). The catalytic activity of ClpR-type ClpP-Isoforms has not been shown so far.

At least one ClpP and two ClpX proteins may be targeted to mitochondria in Arabidopsis as deduced from N-terminal signal sequences. ClpP Proteases are conserved in bacteria. The ClpP protease in E.coli was formerly known as "protease Ti". A knock out

of the protease Ti was shown to be not lethal. E.coli ClpP is assambled as a complex of 14 ClpP subunits in two heptameric rings. Co-immunoprecipitation suggests complexes of similar sizes and an ATP-dependet interaction of ClpP and ClpC subunits in Chloroplasts of Arabidopsis thaliana (Halperin et al. 2001, Planta 213, pp. 614-619). Furthermore, a 350kDa ClpP complex has been identified in Arabidopsis chloroplasts using blue native gel electrophoresis. The complex presumingly containins most of the known ClpP Isoenzymes (Benoit-Peltier et al. 2001, Journal of Biological Chemistry 276, pp. 16348-16327). Consequently the complexity and redundancy of plant Clp proteases is high and detailed information about composition of the clp complex and the functional role of its subunits remain to be clarified. Particularly the role of ClpP redundancy is still unclear.

The ClpP subunit is capable of activly hydrolysing peptides of max. 5 aminoacids in vitro. ClpA,B,C subunits constitute ATP-hydrolysing chaperones which unfold target-proteins and present them for hydrolysis to ClpP (Porankiewicz et al. 1999, Molecular Microbiology 32, 449-458). Involvement of ClpP in the degradation of the cytochrome b6f complex an PSII has been decribed in Chlamydomonas (Majeran et al. 2001, Plant Physiology 23+, pp. 421-433). Functional properties of ClpR-type Clp-Proteases as well as the ClpP like Proeases are yet to be determined.
Surprisingly, it has been found within the scope of the present invention that plants in which a Clp protease was reduced in a selective manner have phenotypes which are comparable with phenotypes generated by herbicide application. Drastic growth retardation and damage such as were observed.

The present invention relates to the use of a polypeptide, which has the activity of nuclear encoded Clp-protease in a method for identifying herbicides, preferably of a polypeptide, which has the activity of nuclear encoded Clp-protease, which is

a) selected from the group consisting of ClpP1-protease, ClpP2-protease, ClpP3-protease, ClpP4-protease and ClpP6-protease; or

b) selected from the group consisting of ClpR1-protease, ClpR3-protease, ClpR4-protease; or

c) ClpP-like-protease, wherein more preferably

a) the ClpP1-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:1, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:2 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:1 which has an identity with SEQ ID NO:1 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:1, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:2;

b) the ClpP2-protease encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:3, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:4 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:3 which has an identity with SEQ ID NO:3 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:3, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:4;

c) the ClpP3-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:5, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:6 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:5 which has an identity with SEQ ID NO:5 of has at least 50%;or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:5, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:6;

d) the ClpP4-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:7, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:8 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:7 which has an identity with SEQ ID NO:7 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:7, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:8;

e) the ClpP6-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:9, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:10 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:9 which has an identity with SEQ ID NO:9 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:9, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:10;

f) the ClpR1-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:11, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:12 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:11 which has an identity with SEQ ID NO:11 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:11, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:12;

g) the ClpR3-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:13, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:14 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:13 which has an identity with SEQ ID NO:13 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:13, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:14;

h) the ClpR4-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:15, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:16 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:15 which has an identity with SEQ ID NO:15 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:15, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:16;

i) the ClpP like-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:17, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:18 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:17 which has an identity with SEQ ID NO:17 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:17, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:18;

wherein the sequences b) i-iv, e) i-iv, f) i-iv and are especially preferred
The term "comprising" in relation to a nucleic acid sequence means that the nucleic acid sequence can be flanked by additional nucleic acid sequences that have on the 5' end and on the 3' end or on the 5'end or on the 3' end on the end a sequence length of at least 1000 bp, preferably at least 500 bp, more preferably at least 250bp, most preferably at least 100bp.

The functional equivalent according to the invention of SEQ ID NO:1 as described in a) iii), which encodes a polypeptide, which has the activity of nuclear encoded Clp-protease, and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:1.

The functional equivalents of the nucleic acid sequence SEQ ID NO:1 set forth in a) iv. are encoded by an amino acid sequence, which has the activity of nuclear encoded Clp-protease and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:2.

The functional equivalent according to the invention of SEQ ID NO:3 as described in b) iii), which encodes a polypeptide, which has the activity of nuclear encoded Clp-protease, and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%,

81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:3.

An example of a functional equivalent of SEQ ID NO: 3 is the nucleic acid sequence of Arabidopsis thaliana (Gene Bank Acc. No. AB022327). This sequence is herein incorporated by reference.

The functional equivalents of the nucleic acid sequence set forth SEQ ID NO:3 in b) iv. are encoded by an amino acid sequence, which has the activity of nuclear encoded Clp-protease and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:4.

The functional equivalent according to the invention of SEQ ID NO:5 as described in c) iii), which encodes a polypeptide, which has the activity of nuclear encoded Clp-protease, and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:5.

The functional equivalents of the nucleic acid sequence SEQ ID NO:5 set forth in c) iv. are encoded by an amino acid sequence, which has the activity of nuclear encoded Clp-protease and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:6.

The functional equivalent according to the invention of SEQ ID NO:7 as described in d) iii), which encodes a polypeptide, which has the activity of nuclear encoded Clp-protease, and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:7.

The functional equivalents of the nucleic acid sequence set forth SEQ ID NO:7 in d) iv. are encoded by an amino acid sequence, which has the activity of nuclear encoded Clp-protease and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%,

57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:7.

The functional equivalent according to the invention of SEQ ID NO:9 as described in e) iii), which encodes a polypeptide, which has the activity of nuclear encoded Clp-protease, and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:9.

The functional equivalents of the nucleic acid sequence set forth SEQ ID NO:9 in e) iv. are encoded by an amino acid sequence, which has the activity of nuclear encoded Clp-protease and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:10.

The functional equivalent according to the invention of SEQ ID NO:11 as described in f) iii), which encodes a polypeptide, which has the activity of nuclear encoded Clp-protease, and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:11.

The functional equivalents of the nucleic acid sequence SEQ ID NO:11 set forth in f) iv. are encoded by an amino acid sequence, which has the activity of nuclear encoded Clp-protease and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:12.

An example of a functional equivalent of SEQ ID NO: 11 is the nucleic acid sequence of Arabidopsis thaliana (Gene Bank Acc. No. AB022330). This sequence is herein incorporated by reference.

The functional equivalent according to the invention of SEQ ID NO:13 as described in g) iii), which encodes a polypeptide, which has the activity of nuclear encoded Clp-protease, and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:13.

The functional equivalents of the nucleic acid sequence SEQ ID NO:13 set forth in g) iv. are encoded by an amino acid sequence, which has the activity of nuclear encoded Clp-protease and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:14.

The functional equivalent according to the invention of SEQ ID NO:15 as described in h) iii), which encodes a polypeptide, which has the activity of nuclear encoded Clp-protease, and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:15.

The functional equivalents of the nucleic acid sequence SEQ ID NO:15 set forth in h) iv. are encoded by an amino acid sequence, which has the activity of nuclear encoded Clp-protease and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:16.

The functional equivalent according to the invention of SEQ ID NO:17 as described in i) iii), which encodes a polypeptide, which has the activity of nuclear encoded Clp-protease, and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:17.

The functional equivalents of the nucleic acid sequence SEQ ID NO:17 set forth in i) iv. are encoded by an amino acid sequence, which has the activity of nuclear encoded

Clp-protease and has at least an identity of 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64% or 65% or preferably of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% or 79% more preferably of 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% or 90% most preferably of 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:18.

An example of a functional equivalent of SEQ ID NO: 17 is the nucleic acid sequence of Arabidopsis thaliana (Gene Bank Acc. No. AK118525). This sequence is herein incorporated by reference.

Furthermore claimed within the scope of the present invention are plant nucleic acid sequence

I) encoding a ClpP2-protease comprising:

a) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:3, or

b) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:4 by backtranslating, or

c) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:1 which has an identity with SEQ ID NO:3 of has at least 66%; or

d) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:11, which is encoded by an amino acid sequence that has at least an identity of 76% with the SEQ ID NO:4;

II) encoding a ClpR1-protease comprising:

a) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:11, or

b) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:12 by backtranslating, or

c) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:1 which has an identity with SEQ ID NO:11 of has at least 69%; or

d) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:11, which is encoded by an amino acid sequence that has at least an identity of 71% with the SEQ ID NO:12;

III) encoding a ClpP-like-protease comprising:

a) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:17, or

b) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:18 by backtranslating, or

c) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:1 which has an identity with SEQ ID NO:17 of has at least 67%; or

d) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:17, which is encoded by an amino acid sequence that has at least an identity of 79% with the SEQ ID NO:18;

The functional equivalent of SEQ ID NO:3 set forth in I c) has at least an identity of 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, by preference at least 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82% or 83%, preferably at least 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92% or 93%, especially preferably at least 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:3.

The functional equivalents of the nucleic acid sequence SEQ ID NO:3 set forth in I) d) are encoded by an amino acid sequence, which has the activity of nuclear encoded Clp-protease and has at least an identity of 77%, by preference at least 78%, 79%, 80%, 81%, 82% or 83%, preferably at least 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, especially preferably at least 94%, 95%, 96%, 97%, 98%, 99% with SEQ ID NO:4.

The functional equivalent of SEQ ID NO:11 set forth in II c) has at least an identity of 69%, 70%, 71%, 72%, 73% or 74%, by preference at least 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82% or 83%, preferably at least 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92% or 93%, especially preferably at least 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO:11.

The functional equivalents of the nucleic acid sequence SEQ ID NO:11 set forth in II) d) are encoded by an amino acid sequence, which has the activity of nuclear encoded Clp-protease and has at least an identity of 71% by preference at least 72%, 73%,74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, preferably at least 84%, 85%, 86%,

87%, 88%, 89%, 90%, 91%, 92%, 93%, especially preferably at least 94%, 95%, 96%, 97%, 98%, 99% with SEQ ID NO:12.

The functional equivalent of SEQ ID NO:17 set forth in I c) has at least an identity of 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, by preference at least 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82% or 83%, preferably at least 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92% or 93%, especially preferably at least 94%, 95%, 96%, 97%, 98% or 99% with SEQ ID NO: 17.

The functional equivalents of the nucleic acid sequence SEQ ID NO:17 set forth in I) d) are encoded by an amino acid sequence, which has the activity of nuclear encoded Clp-protease and has at least an identity of 79%, by preference at least 79%, 80%, 81%, 82% or 83%, preferably at least 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, especially preferably at least 94%, 95%, 96%, 97%, 98%, 99% with SEQ ID NO:18.

The polypeptides encoded by the abovementioned nucleic acid sequences according to I c)-d), II c)-d) and III c)-d) are likewise claimed. The functional equivalents as described in c) and d) are distinguished by the same functionality, i.e. they have the activity of a clp-protease.

The nucleic acid sequences I c)-d), II c)-d) and III c)-d) are hereinbelow termed NCLP-sequences.

The term "nucleic acid sequences according to the invention" which is used hereinbelow refers to nucleic acid sequences encoding a polypeptide, which has the activity of nuclear encoded Clp-protease in a method for identifying herbicides, preferably of a polypeptide, which has the activity of nuclear encoded Clp-protease, which is

a) selected from the group consisting of ClpP1-protease, ClpP2-protease, ClpP3-protease, ClpP4-protease and ClpP6-protease; or

b) selected from the group consisting of ClpR1-protease, ClpR3-protease, ClpR4-protease; or

c) ClpP-like-protease, wherein more preferably

a) the ClpP1-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:1, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:2 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:1 which has an identity with SEQ ID NO:1 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:1, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:2;

b) the ClpP2-protease encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:3, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:4 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:3 which has an identity with SEQ ID NO:3 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:3, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:4;

c) the ClpP3-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:5, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:6 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:5 which has an identity with SEQ ID NO:5 of has at least 50%;or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:5, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:6;

d) the ClpP4-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:7, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:8 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:7 which has an identity with SEQ ID NO:7 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:7, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:8;

e) the ClpP6-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:9, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:10 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:9 which has an identity with SEQ ID NO:9 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:9, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:10;

f) the ClpR1-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:11, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:12 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:11 which has an identity with SEQ ID NO:11 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:11, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:12;

g) the ClpR3-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:13, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:14 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:13 which has an identity with SEQ ID NO:13 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:13, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:14;

h) the ClpR4-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:15, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:16 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:15 which has an identity with SEQ ID NO:15 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:15, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:16;

i) the ClpP like-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:17, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:18 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:17 which has an identity with SEQ ID NO:17 of has at least 50%; or

iv) a functional equivalent of the nucleic acid sequence shown in SEQ ID NO:17, which is encoded by an amino acid sequence that has at least an identity of 50% with the SEQ ID NO:18;

wherein the sequences b) i-iv, e) i-iv, f) i-iv and are especially preferred

A polypeptide, which has the activity of nuclear encoded Clp-protease and is encoded by a nucleic acid sequence according to the invention are hereinbelow simply referred to as "CLP".

Reduced amounts of glyoxysomal CLP cause growth retardation and necrotic and chlorotic leaves in plants.

The gene products of the nucleic acids according to the invention constitute novel targets for herbicides, which make possible the provision of novel herbicides for controlling undesired plants. Moreover, the gene products of the nucleic acids according to the invention constitute novel targets for growth regulators which make possible the provision of novel growth regulators for regulating the growth of plants.

Undesired plants are understood as meaning, in the broadest sense, all those plants which grow at locations where they are undesired, for example:

Dicotyledonous weeds of the genera: Sinapis, Lepidium, Galium, Stellaria, Matricaria, Anthemis, Galinsoga, Chenopodium, Urtica, Senecio, Amaranthus, Portulaca, Xanthium, Convolvulus, Ipomoea, Polygonum, Sesbania, Ambrosia, Cirsium, Carduus, Sonchus, Solanum, Rorippa, Rotala, Lindernia, Lamium, Veronica, Abutilon, Emex, Datura, Viola, Galeopsis, Papaver, Centaurea, Trifolium, Ranunculus, Taraxacum.

Monocotyledonous weeds from the genera: Echinochloa, Setaria, Panicum, Digitaria, Phleum, Poa, Festuca, Eleusine, Brachiaria, Lolium, Bromus, Avena, Cyperus, Sorghum, Agropyron, Cynodon, Monochoria, Fimbristylis, Sagittaria, Eleocharis, Scirpus, Paspalum, Ischaemum, Sphenoclea, Dactyloctenium, Agrostis, Alopecurus, Apera.

SEQ ID NO:1; 3, 5, 7, 9, 11, 13, 15, 17 or 19-21 or parts of SEQ ID NO: 1; 3, 5, 7, 9, 11, 13, 15, 17 or 19-21 can be used for the preparation of hybridization probes. The preparation of these probes and the experimental procedure is known. For example, this can be effected via the selective preparation of radioactive or nonradioactive probes by PCR and the use of suitably labeled oligonucleotides, followed by hybridization experiments. The technologies required for this purpose are detailed, for example, in T. Maniatis, E.F. Fritsch and J. Sambrook, "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989). The probes in question can furthermore be modified by standard technologies (Lit. SDM or random mutagenesis) in such a way that they can be employed for further purposes, for example as a probe which hybridizes specifically with mRNA and the corresponding coding sequences in order to analyze the corresponding sequences in other organisms.

The abovementioned probes can be used for the detection and isolation of functional equivalents of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16 or 18 from other plant species on the basis of sequence identities. In this context, part or all of the sequence of the SEQ ID NO:2 in question is used as a probe for screening a genomic or cDNA library of the plant species in question or in a computer search for sequences of functional equivalents in electronic databases.

Preferred plant species are the undesired plants which have already been mentioned at the outset.

The invention furthermore relates to expression cassettes comprising

a) genetic control sequences in operable linkage with a NCLP sequence; or

b) additional functional elements, or

c) a combination of a) and b);

and to the use of expression cassettes comprising

a) genetic control sequences in operable linkage with a nucleic acid sequence according to the invention,

b) additional functional elements, or

c) a combination of a) and b);

for expressing a CLP, which can be used in in vitro assay systems. Both embodiments of the above-described expression cassettes are referred in the following text as expression cassette according to the invention.

In a preferred embodiment, an expression cassette according to the invention comprises a promoter at the 5' end of the coding sequence and, at the 3' end, a transcription termination signal and, if appropriate, further genetic control sequences which are linked operably with the interposed nucleic acid sequence according to the invention.

The expression cassettes according to the invention are also understood as meaning analogs which can be brought about, for example, by a combination of the individual nucleic acid sequences on a polynucleotide (multiple constructs), on a plurality of polynucleotides in a cell (cotransformation) or by sequential transformation.

Advantageous genetic control sequences under point a) for the expression cassettes according to the invention or for vectors comprising expression cassettes according to the invention are, for example, promoters such as the cos, tac, trp, tet, lpp, lac, laclq, T7, T5, T3, gal, trc, ara, SP6, □-PR or the □-PL promoter, all of which can be used for expressing a CLP, in Gram-negative bacterial strains.

Examples of further advantageous genetic control sequences are present, for example, in the promoters amy and SPO2, both of which can be used for expressing a CLP, in Gram-positive bacterial strains, and in the yeast or fungal promoters AUG1, GPD-1, PX6, TEF, CUP1, PGK, GAP1, TPI, PHO5, AOX1, GAL10/CYC1, CYC1, OliC, ADH, TDH, Kex2, MFA or NMT or combinations of the abovementioned promoters (Degryse et al., Yeast 1995 June 15; 11(7):629-40; Romanos et al. Yeast 1992 June;8(6):423-88; Benito et al. Eur. J. Plant Pathol. 104, 207-220 (1998); Cregg et al. Biotechnology (N Y) 1993 Aug;11(8):905-10; Luo X., Gene 1995 Sep 22;163(1):127-31: Nacken et al., Gene 1996 Oct 10;175(1-2): 253-60; Turgeon et al., Mol Cell Biol 1987 Sep;7(9):3297-305) or the transcription terminators NMT, Gcy1, TrpC, AOX1, nos, PGK or CYC1 (Degryse et al., Yeast 1995 June 15; 11(7):629-40; Brunelli et al. Yeast 1993 (Dec9(12): 1309-18; Frisch et al., Plant Mol. Biol. 27(2), 405-409 (1995); Scorer et al., Biotechnology (N.Y. 12 (2), 181-184 (1994), Genbank acc. number Z46232; Zhao et al. Genbank acc number : AF049064; Punt et al., (1987) Gene 56 (1), 117-124), all of which can be used for expressing CLP, in yeast strains.

Examples of genetic control sequences which are suitable for expression in insect cells are the polyhedrin promoter and the p10 promoter (Luckow, V.A. and Summers, M.D. (1988) Bio/Techn. 6, 47-55).

Advantageous genetic control sequences for expressing CLP, in cell culture, in addition to polyadenylation sequences such as, for example, from simian virus 40, are eu-

karyotic promoters of viral origin such as, for example, promoters of the polyoma virus, adenovirus 2, cytomegalovirus or simian virus 40.

Further advantageous genetic control sequences for expressing nuclear encoded Clp Protease, in plants are present in the plant promoters CaMV/35S [Franck et al., Cell 21(1980) 285-294], PRP1 [Ward et al., Plant. Mol. Biol. 22 (1993)], SSU, OCS, LEB4, USP, STLS1, B33, NOS; FBPaseP (WO 98/18940) or in the ubiquitin or phaseolin promoter; a promoter which is preferably used being, in particular, a plant promoter or a promoter derived from a plant virus. Especially preferred are promoters of viral origin such as the promoter of the cauliflower mosaic virus 35S transcript (Franck et al., Cell 21 (1980), 285-294; Odell et al., Nature 313 (1985), 810-812). Further preferred constitutive promoters are, for example, the agrobacterium nopaline synthase promoter, the TR double promoter, the agrobacterium OCS (octopine synthase) promoter, the ubiquitin promoter, (Holtorf S et al., Plant Mol Biol 1995, 29:637-649), the promoters of the vacuolar ATPase subunits, or the promoter of a proline-rich wheat protein (WO 91/13991).

The expression cassettes may also comprise, as genetic control sequence, a chemically inducible promoter, by which the expression of the exogenous gene in the plant can be controlled at a specific point in time. Such promoters, such as, for example, the PRP1 promoter (Ward et al., Plant. Mol. Biol. 22 (1993), 361-366), a salicylic-acid-inducible promoter (WO 95/19443), a benzenesulfonamide-inducible promoter (EP-A-0388186), a tetracyclin-inducible promoter (Gatz et al., (1992) Plant J. 2, 397404), an abscisic-acid-inducible promoter (EP-A 335528) or an ethanol- or cyclohexanone-inducible promoter (WO 93/21334) may also be used.

Furthermore, suitable promoters are those which confer tissue- or organ-specific expression in, for example, anthers, ovaries, flowers and floral organs, leaves, stomata, trichomes, stems, vascular tissues, roots and seeds. Others which are suitable in addition to the abovementioned constitutive promoters are, in particular, those promoters which ensure leaf-specific expression. Promoters which must be mentioned are the potato cytosolic FBPase promoter (WO 97/05900), the rubisco (ribulose-1,5-bisphosphate carboxylase) SSU (small subunit) promoter or the ST-LSI promoter from potato (Stockhaus et al., EMBO J. 8 (1989), 2445 - 245). Promoters which are furthermore preferred are those which control expression in seeds and plant embryos. Examples of seed-specific promoters are the phaseolin promoter (US 5,504,200, Bustos MM et al., Plant Cell. 1989;1(9):839-53), the promoter of the 2S albumin gene (Joseffson LG et al., J Biol Chem 1987, 262:12196-12201), the legumin promoter (Shirsat A et al., Mol Gen Genet. 1989;215(2):326-331), the USP (unknown seed protein) promoter (Bäumlein H et al., Molecular & General Genetics 1991, 225(3):459-67), the napin gene promoter (Stalberg K, et al., L. Planta 1996, 199:515-519), the sucrose binding

protein promoter (WO 00/26388) or the LeB4 promoter (Bäumlein H et al., Mol Gen Genet 1991, 225: 121-128; Fiedler, U. et al., Biotechnology (NY) (1995), 13 (10) 1090).

Further promoters which are suitable as genetic control sequences are, for example, specific promoters for tubers, storage roots or roots, such as, for example, the class I patatin promoter (B33), the potato cathepsin D inhibitor promoter, the starch synthase (GBSS1) promoter or the sporamin promoter, fruit-specific promoters such as, for example, the fruit-specific promoter from tomato (EP-A 409625), fruit-maturation-specific promoters such as, for example, the fruit-maturation-specific promoter from tomato (WO 94/21794), inflorescence-specific promoters such as, for example, the phytoene synthase promoter (WO 92/16635) or the promoter of the P-rr gene (WO 98/22593), or plastid- or chromoplast-specific promoters such as, for example, the RNA polymerase promoter (WO 97/06250), or else the Glycine max phosphoribosyl-pyrophosphate amidotransferase promoter (see also Genbank Accession No. U87999), or another node-specific promoter as described in EP-A 249676.

Additional functional elements b) are understood as meaning, by way of example but not by limitation, reporter genes, replication origins, selection markers and what are known as affinity tags, in fusion with CLP, directly or by means of a linker optionally comprising a protease cleavage site. Further suitable additional functional elements are sequences which ensure that the product is targeted into the apoplasts, into plastids, the vacuole, the mitochondrion, the peroxisome, the endoplasmatic reticulum (ER) or, owing to the absence of such operative sequences, remains in the compartment where it is formed, the cytosol, (Kermode, Crit. Rev. Plant Sci. 15, 4 (1996), 285-423).

Also in accordance with the invention are vectors comprising at least one copy of the nucleic acid sequences according to the invention and/or the expression cassettes according to the invention.

In addition to plasmids, vectors are furthermore also understood as meaning all of the other known vectors with which the skilled worker is familiar, such as, for example, phages, viruses such as SV40, CMV, baculovirus, adenovirus, transposons, IS elements, phasmids, phagemids, cosmids or linear or circular DNA. These vectors can be replicated autonomously in the host organism or replicated chromosomally; chromosomal replication is preferred.

In a further embodiment of the vector, the nucleic acid construct according to the invention can advantageously also be introduced into the organisms in the form of a linear DNA and integrated into the genome of the host organism via heterologous or homologous recombination. This linear DNA may consist of a linearized plasmid or only of the nucleic acid construct as vector, or the nucleic acid sequences used.

Further prokaryotic or eukaryotic expression systems are mentioned in Chapters 16 and 17 in Sambrook et al., "Molecular Cloning: A Laboratory Manual." 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. Further advantageous vectors are described in Hellens et al. (Trends in plant science, 5, 2000).

The expression cassette according to the invention and vectors derived therefrom can be used for transforming bacteria, cyanobacteria, (for example of the genus Synechocystis, Anabaena, Calothrix, Scytonema, Oscillatoria, Plectonema and Nostoc), proteobacteria such as, for example, Magnetococcus sp. MC1, yeasts, filamentous fungi and algae and eukaryoatic nonhuman cells (for example insect cells) with the aim of producing CLP, recombinantly, the generation of a suitable expression cassette depending on the organism in which the gene is to be expressed.

Vectors comprising a NCLP sequence form part of the subject-matter of the present invention.

In a further advantageous embodiment, the nucleic acid sequences according to the invention may also be introduced into an organism by themselves.

If, in addition to the nucleic acid sequences, further genes are to be introduced into the organism, they can all be introduced into the organism together in a single vector, or each individual gene can be introduced into the organism in each case in one vector, it being possible to introduce the different vectors simultaneously or in succession.

In this context, the introduction, into the organisms in question (transformation), of the nucleic acid(s) according to the invention, of the expression cassette or of the vector can be effected in principle by all methods with which the skilled worker is familiar.

In the case of microorganisms, the skilled worker will find suitable methods in the textbooks by Sambrook, J. et al. (1989) "Molecular cloning: A laboratory manual", Cold Spring Harbor Laboratory Press, von F.M. Ausubel et al. (1994) "Current protocols in molecular biology", John Wiley and Sons, by D.M. Glover et al., DNA Cloning Vol.1, (1995), IRL Press (ISBN 019-963476-9), by Kaiser et al. (1994) Methods in Yeast Genetics, Cold Spring Habor Laboratory Press or Guthrie et al. "Guide to Yeast Genetics and Molecular Biology", Methods in Enzymology, 1994, Academic Press. In the transformation of filamentous fungi, the methods of choice are firstly the generation of protoplasts and transformation with the aid of PEG (Wiebe et al. (1997) Mycol. Res. 101 (7): 971-877; Proctor et al. (1997) Microbiol. 143, 2538-2591), and secondly the transformation with the aid of Agrobacterium tumefaciens (de Groot et al. (1998) Nat. Biotech. 16, 839-842).

In the case of dicots, the methods which have been described for the transformation and regeneration of plants from plant tissues or plant cells can be exploited for transient or stable transformation. Suitable methods are the biolistic method or the transformation of protoplasts (cf., for example, Willmitzer, L., 1993 Transgenic plants. In: Biotechnology, A Multi-Volume Comprehensive Treatise (H.J. Rehm, G. Reed, A. Pühler, P. Stadler, eds.), Vol. 2, 627-659, VCH Weinheim-New York-Basle-Cambridge), electroporation, the incubation of dry embryos in DNA-containing solution, microinjection and the agrobacterium-radiated gene transfer. The abovementioned methods are described, for example, in B. Jenes et al., Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, edited by S.D. Kung and R. Wu, Academic Press (1993) 128-143 and in Potrykus, Annu. Rev. Plant Physiol. Plant Molec.Biol. 42 (1991) 205-225).

The transformation by means of agrobacteria, and the vectors to be used for the transformation, are known to the skilled worker and described extensively in the literature (Bevan et al., Nucl. Acids Res. 12 (1984) 8711. The intermediary vectors can be integrated into the agrobacterial Ti or Ri plasmid by means of homologous recombination owing to sequences which are homologous to sequences in the T-DNA. This plasmid additionally contains the vir region, which is required for the transfer of the T-DNA. Intermediary vectors are not capable of replication in agrobacteria. The intermediary vector can be transferred to Agrobacterium tumefaciens by means of a helper plasmid (conjugation). Binary vectors are capable of replication both in E. coli and in agrobacteria. They contain a selection marker gene and a linker or polylinker which are framed by the right and left T-DNA border region. They can be transformed directly into the agrobacteria (Holsters et al. Mol. Gen. Genet. 163 (1978), 181-187), EP A 0 120 516; Hoekema, in: The Binary Plant Vector System Offsetdrukkerij Kanters B.V., Alblasserdam (1985), Chapter V; Fraley et al., Crit. Rev. Plant. Sci., 4: 1-46 and An et al. EMBO J. 4 (1985), 277-287).

The transformation of monocots by means of vectors based on agrobacterium has also been described (Chan et al., Plant Mol. Biol. 22(1993), 491-506; Hiei et al., Plant J. 6 (1994) 271-282; Deng et al. Science in China 33 (1990), 28-34; Wilmink et al., Plant Cell Reports 11,(1992) 76-80; May et al. Biotechnology 13 (1995) 486-492; Conner and Domisse; Int. J. Plant Sci. 153 (1992) 550-555; Ritchie et al. Transgenic Res. (1993) 252-265). Alternative systems for the transformation of monocots are the transformation by means of biolistic approach (Wan and Lemaux; Plant Physiol. 104 (1994), 37-48; Vasil et al. Biotechnology 11 (1992), 667-674; Ritala et al., Plant Mol. Biol 24, (1994) 317-325; Spencer et al., Theor. Appl. Genet. 79 (1990), 625-631), protoplast transformation, the electroporation of partially permeabilized cells, and the introduction of DNA by means of glass fibers. In particular the transformation of maize has been described repeatedly in the literature (cf., for example, WO 95/06128; EP 0513849 A1; EP 0465875 A1; EP 0292435 A1; Fromm et al., Biotechnology 8 (1990), 833-844;

Gordon-Kamm et al., Plant Cell 2 (1990), 603-618; Koziel et al., Biotechnology 11(1993) 194-200; Moroc et al., Theor Applied Genetics 80 (190) 721-726).

The successful transformation of other cereal species has also already been described for example in the case of barley (Wan and Lemaux, see above; Ritala et al., see above; wheat (Nehra et al., Plant J. 5(1994) 285-297).

Agrobacteria which have been transformed with a vector according to the invention can likewise be used in a known manner for the transformation of plants, such as test plants like Arabidopsis or crop plants like cereals, maize, oats, rye, barley, wheat, soya, rice, cotton, sugarbeet, canola, sunflower, flax, hemp, potato, tobacco, tomato, carrot, capsicum, oilseed rape, tapioca, cassava, arrowroot, Tagetes, alfalfa, lettuce and the various tree, nut and grapevine species, for example by bathing scarified leaves or leaf segments in an agrobacterial solution and subsequently growing them in suitable media.

The genetically modified plant cells can be regenerated via all methods with which the skilled worker is familiar. Such methods can be found in the abovementioned publications by S.D. Kung and R. Wu, Potrykus or Höfgen and Willmitzer.

The transgenic organisms generated by transformation with one of the above-described embodiments of an expression cassette comprising a nucleic acid sequence according to the invention or a vector comprising the abovementioned expression cassette, and the recombinant CLP, which can be obtained from the transgenic organism by means of expression, form part of the subject matter of the present invention. The use of transgenic organisms comprising an expression cassette according to the invention, for example for providing recombinant protein, and/or the use of these organisms in in-vivo assay systems likewise form part of the subject matter of the present invention.

Preferred organisms for the recombinant expression are not only bacteria, yeasts, mosses, algae and fungi, but also eukaryotic cell lines.

Preferred mosses are Physcomitrella patens or other mosses described in Kryptogamen [Cryptogamia], Vol.2, Moose, Farne [Mosses, Ferns], 1991, Springer Verlag (ISBN 3540536515).

Preferred within the bacteria are, for example, bacteria from the genus Escherichia, Erwinia, Flavobacterium, Alcaligenes or cyanobacteria, for example from the genus Synechocystis, Anabaena, Calothrix, Scytonema, Oscillatoria, Plectonema and Nostoc, especially preferably Synechocystis or Anabaena.

Preferred yeasts are Candida, Saccharomyces, Schizosaccheromyces, Hansenula or Pichia.

Preferred fungi are Aspergillus, Trichoderma, Ashbya, Neurospora, Fusarium, Beauveria, Mortierella, Saprolegnia, Pythium, or other fungi described in Indian Chem Engr. Section B. Vol 37, No 1,2 (1995).

Preferred plants are selected in particular among monocotyledonous crop plants such as, for example, cereal species such as wheat, barley, sorghum or millet, rye, triticale, maize, rice or oats, and sugarcane. The transgenic plants according to the invention are, furthermore, in particular selected from among dicotyledonous crop plants such as, for example, Brassicaceae such as oilseed rape, cress, Arabidopsis, cabbages or canola; Leguminosae such as soyabean, alfalfa, pea, beans or peanut, Solanaceae such as potato, tobacco, tomato, egg plant or capsicum; Asteraceae such as sunflower, Tagetes, lettuce or Calendula; Cucurbitaceae such as melon, pumpkin/squash or zucchini, or linseed, cotton, hemp, flax, red pepper, carrot, sugar beet, or various tree, nut and grapevine species.

In principle, transgenic animals such as, for example, C. elegans, are also suitable as host organisms.

Also preferred is the use of expression systems and vectors which are available to the public or commercially available.

Those which must be mentioned for use in E. coli bacteria are the typical advantageous commercially available fusion and expression vectors pGEX [Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) Gene 67:31-40], pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ), which contains glutathione S transferase (GST), maltose binding protein or protein A, the pTrc vectors (Amann et al., (1988) Gene 69:301-315), "pKK233-2" from CLONTECH, Palo Alto, CA and the "pET", and the "pBAD" vector series from Stratagene, La Jolla and the TOPO-TA vector series drom Invitrogen.

Further advantageous vectors for use in yeast are pYepSec1 (Baldari, et al., (1987) Embo J. 6:229-234), pMFa (Kurjan and Herskowitz, (1982) Cell 30:933-943), pJRY88 (Schultz et al., (1987) Gene 54:113-123), and pYES derivatives, pGAPZ derivatives, pPICZ derivatives, and the vectors of the "Pichia Expression Kit" (Invitrogen Corporation, San Diego, CA). Vectors for use in filamentous fungi are described in: van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, J.F. Peberdy, et al., eds., p. 1-28, Cambridge University Press: Cambridge.

As an alternative, insect cell expression vectors may also be used advantageously, for example for expression in Sf9, Sf21 or Hi5 cells, which are infected via recombinant Baculoviruses. Examples of these are the vectors of the pAc series (Smith et al. (1983) Mol. Cell Biol. 3:2156-2165) and the pVL series (Lucklow and Summers (1989) Virology 170:31-39). Others which may be mentioned are the Baculovirus expression systems "MaxBac 2.0 Kit" and "Insect Select System" from Invitrogen, Carlsbad or "BacPAK Baculovirus Expression System" from CLONTECH, Palo Alto, CA. Insect cells are particularly suitable for overexpressing eukaryotic proteins since they effect posttranslational modifications of the proteins which are not possible in bacteria and yeasts. The skilled worker is familiar with the handling of cultured insect cells and with their infection for expressing proteins, which can be carried out analogously to known methods (Luckow and Summers, Bio/Tech. 6, 1988, pp.47-55; Glover and Hames (eds) in DNA Cloning 2, A practical Approach, Expression Systems, Second Edition, Oxford University Press, 1995, 205-244).

Plant cells or algal cells are others which can be used advantageously for expressing genes. Examples of plant expression vectors can be found as mentioned above in Becker, D., et al. (1992) "New plant binary vectors with selectable markers located proximal to the left border", Plant Mol. Biol. 20: 1195-1197 or in Bevan, M.W. (1984) "Binary Agrobacterium vectors for plant transformation", Nucl. Acid. Res. 12: 8711-8721.

Moreover, the nucleic acid sequences according to the invention can be expressed in mammalian cells. Examples of suitable expression vectors are pCDM8 and pMT2PC, which are mentioned in: Seed, B. (1987) Nature 329:840 or Kaufman et al. (1987) EMBO J. 6:187-195). Promoters preferably to be used in this context are of viral origin such as, for example, promoters of polyoma virus, adenovirus 2, cytomegalovirus or simian virus 40. Further prokaryotic and eukaryotic expression systems are mentioned in Chapter 16 and 17 in Sambrook et al., Molecular Cloning: A Laboratory Manual. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. Further advantageous vectors are described in Hellens et al. (Trends in plant science, 5, 2000).

The transgenic organisms which comprise a NCLP sequence are claimed within the scope of the present invention.

All of the above-described embodiments of the transgenic organisms, which comprise at least one nucleic acid sequence according to the invention come under the term "transgenic organism according to the invention".

The present invention furthermore relates to the use of CLP, in a method for identifying herbicidally active test compounds.

The method according to the invention for identifying herbicidally active compounds preferably comprises the following steps:

i. bringing CLP into contact with one or more test compounds under conditions which permit the test compound(s) to bind to a nucleic acid sequence according to the invention or to CLP, and

ii. detecting whether the test compound binds to the CLP of i), or

iii. detecting whether the test compound reduces or blocks the enzymatic or biological activity of CLP of i), or

iv. detecting whether the test compound reduces or blocks the transcription, translation or expression of CLP of i).

The detection in accordance with step (ii) of the above method can be effected using techniques which identify the interaction between the polypeptide and ligand. In this context, either the test compound or the enzyme can contain a detectable label such as, for example, a fluorescent label, a radioisotope, a chemiluminescent label or an enzyme label. Examples of enzyme labels are horseradish peroxidase, alkaline phosphatase or luciferase. The subsequent detection depends on the label and is known to the skilled worker.

In this context, five preferred embodiments which are also suitable for high-throughput methods (HTS) in connection with the present invention must be mentioned in particular:

1. The average diffusion rate of a fluorescent molecule as a function of the mass can be determined in a small sample volume via fluorescence correlation spectroscopy (FCS) (Proc. Natl. Acad. Sci. USA (1994) 11753-11575). FCS can be employed for determining protein/ligand interactions by measuring the change in the mass, or the changed diffusion rate which this entails, of a test compound when binding to CLP. A method according to the invention can be designed directly for measuring the binding of a test compound labeled by a fluorescent molecule. As an alternative, the method according to the invention can be designed in such a way that a chemical reference compound which is labeled by a fluorescent molecule is displaced by further test compounds ("displacement assay").

2. Fluoresence polarization exploits the characteristic of a quiescent fluorophore excited with polarized light to likewise emit polarized light. If, however, the fluoro-

phore is allowed to rotate during the excited state, the polarization of the fluorescent light which is emitted is more or less lost. Under otherwise identical conditions (for example temperature, viscosity, solvent), the rotation is a function of molecule size, whereby findings regarding the size of the fluorophore-bound residue can be obtained via the reading (Methods in Enzymology 246 (1995), pp. 283-300). A method according to the invention can be designed directly for measuring the binding of a test compound labeled with a fluorescent molecule to the CLP. As an alternative, the method according to the invention may also take the form of the "displacement assay" described under 1.

3. Fluorescence resonance energy transfer (FRET) is based on the irradiation-free energy transfer between two spatially adjacent fluorescent molecules under suitable conditions. A prerequisite is that the emission spectrum of the donor molecule overlaps with the excitation spectrum of the acceptor molecule. The fluorescent label of CLP, and binding test compound, the binding can be measured by means of FRET (Cytometry 34, 1998, pp. 159-179). As an alternative, the method according to the invention may also take the form of the "displacement assay" described under 1. An especially suitable embodiment of FRET technology is "Homogeneous Time Resolved Fluorescence" (HTRF) as can be obtained from Packard BioScience.

4. Surface-enhanced laser desorption/ionization (SELDI) in combination with a time-of-flight mass spectrometer (MALDI-TOF) makes possible the rapid analysis of molecules on a support and can be used for analyzing protein/ligand interactions (Worral et al., (1998) Anal. Biochem. 70:750-756). In a preferred embodiment, CLP, is immobilized on a suitable support and incubated with the test compound. After one or more suitable wash steps, the test compound molecules which are additionally bound to CLP, can be detected by means of the abovementioned methodology and test compounds which are bound to CLP, can thus be selected.

5. The measurement of surface plasmon resonance is based on the change in the refractive index at a surface when a test compound binds to a protein which is immobilized to said surface. Since the change in the refractive index is identical for virtually all proteins and polypeptides for a defined change in the mass concentration at the surface, this method can be applied to any protein in principle (Lindberg et al. Sensor Actuators 4 (1983) 299-304; Malmquist Nature 361 (1993) 186-187). The measurement can be carried out for example with the automatic analyzer based on surface plasmon resonance which is available from Biacore (Freiburg) at a throughput of, currently, up to 384 samples per day. A method according to the invention can be designed directly for measuring the binding of a test compound to CLP. As an alternative, the method according to

the invention may also take the form of the "displacement assay" described under 1.

The compounds identified via the abovementioned methods 1 to 5 may be suitable as inhibitors. All of the substances identified via the abovementioned methods can subsequently be checked for their herbicidal action in another embodiment of the method according to the invention.

Furthermore, there exists the possibility of detecting further candidates for herbicidal active ingredients by molecular modeling via elucidation of the three-dimensional structure of CLP, by x-ray structure analysis. The preparation of protein crystals required for x-ray structure analysis, and the relevant measurements and subsequent evaluations of these measurements, the detection of a binding site in the protein, and the prediction of potential inhibitor structures are known to the skilled worker. In principle, an optimization of the compound identified by the abovementioned methods is also possible via molecular modeling.

A preferred embodiment of the method according to the invention, which is based on steps i) and ii), consists in selecting a test compound which reduces or blocks the activity of the CLP. Preferably, the activity of the CLP, incubated with the test compound is herein compared with the activity of a CLP, not incubated with a test compound.

A more preferred embodiment of the method based on steps i) and ii) consists in

i. expressing CLP in a transgenic organism according to the invention or growing an organism which naturally contains a CLP,

ii. bringing CLP, of step i) in the cell digest of the transgenic or nontransgenic organism, in partially purified or in homogeneously purified form, into contact with a test compound; and

iii. selecting a compound which reduces or blocks the activity of the nuclear encoded Clp Protease. Preferably the activity of CLP incubated with the test compound is herein compared with the activity of a CLP, not incubated with a test compound.

The solution containing the CLP, can consist of the lysate of the original organism or of the transgenic organism which has been transformed with an expression cassette according to the invention. If necessary, the CLP, can be purified partially or fully via customary methods. A general overview over current protein purification techniques is described, for example, in Ausubel, F.M. et al., Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley-Interscience (1994); ISBN 0-87969-309-6. In the

case of recombinant preparation, the protein which has been fused with an affinity tag can be purified via affinity chromatography as is known to the skilled worker.

The CLP, which is required for in vitro methods can thus be isolated either by means of heterologous expression from a transgenic organism according to the invention or from an organism containing CLP, for example from an undesired plant, the term "undesired plant" being understood as meaning the species mentioned at the outset.

To identify herbicidal compounds, the CLP, is now incubated with a test compound. After a reaction time, the enzymatic activity of the CLP, incubated with the test compound is determined in comparison with a CLP, not incubated with a test compound. If the CLP, is inhibited, a significant decrease in activity in comparison with the activity of the noninhibited polypeptide according to the invention is observed, the result being a reduction of at least 10%, advantageously at least 20%, preferably at least 30%, especially preferably by at least 50%, up to 100% reduction (blocking). Preferred is an inhibition of at least 50% at test compound concentrations of 10^{-4}M, preferably at 10^{-5}M, especially preferably of 10^{-6}M, based on enzyme concentration in the micromolar range.

The enzymatic activity of CLP, can be determined for example by an activity assay in which the increase of the product, the decrease of the substrate (or starting material) or the decrease or increase of the cofactor are determined, or by a combination of at least two of the abovementioned parameters, as a function of a defined period of time.

Examples of suitable substrates are, for example small peptides and modified small peptides like peptides coupled to a fluorogenic molecule such as aminomethylcoumarin and succinylated peptides.

If appropriate, derivatives of the abovementioned compounds which contain a detectable label such as, for example, a fluorescent label (e.g. fluorogenic substrates such as N-Suc-Leu-Tyr-(7-amino-4-methylcoumarine) (SLT-AMC), Suc-Ala-Ala-Ala-AMC, Suc-Leu-Leu-Val-Tyr-AMC, Suc-Ala-Ala-Phe-AMC, Suc-Ile-Ile-Trp-AMC, Suc-Ala-Phe-Lys-AMC), a radioisotope label or a chemiluminescent label, may also be used.

The amounts of substrates to be employed in the activity tests may range between 0.5 and 100 mM, based on 1-100 µg/ml enzyme.

[

The activity can be determined for example by tracking Proteolysis fluorimetrically when using fluorogenic Peptide substrates analogously to the method described by Woo et al. 1989 The Journal of Biological Chemistry 264, pp.2088-2091, which is herein incorporated by reference.

The activity may also be determined in an ATP-dependent fashion in the presence of ClpA, ClpB or ClpC Protein as described in Halperin et al. 2001, Planta 213, pp. 614-619. The preferred Substrate is then b-casein.

Furthermore the activity may be measured by HPLC and HPLC-MS mehtods detecting fragments of the peptides used as substrates.

Another preferred embodiment of the method according to the invention which is based on steps i) and iii) consists of the following steps:

i. generating a transgenic organism according to the invention comprising a nucleic acid sequence according to the invention, wherein CLP is expressed recombinantly;

ii. applying a test compound to the transgenic organism of i) and to a nontransgenic organism of the same species;

iii. determining the growth or the viability of the transgenic and the nontransgenic organisms after application of the test substance, and

iv. selecting test compounds which bring about a reduced growth or a limited viability of the nontransgenic organism in comparison with the growth of the transgenic organism.

In this context, the difference in growth in step iv) for the selection of a herbicidally active inhibitor amounts to at least 10%, by preference 20%, preferably 30%, especially preferably 40% and very especially preferably 50%.

The transgenic organism in this context is preferably a plant, an alga, a cyanobacterium, for example of the genus Synechocystis or a proteobacterium such as, for example, Magnetococcus sp. MC1, preferably plants which can be transformed by means of customary techniques, such as Arabidopsis thaliana Allium cepa, Ananas comosus, Arachis hypogaea, Asparagus officinalis, Beta vulgaris spec. altissima, Beta vulgaris spec. rapa, Brassica napus var. napus, Brassica napus var. napobrassica, Brassica rapa var. silvestris, Camellia sinensis, Carthamus tinctorius, Carya illinoinensis, Citrus limon, Citrus sinensis, Coffea arabica (Coffea canephora, Coffea liberica), Cucumis sativus, Cynodon dactylon, Daucus carota, Elaeis guineensis, Fragaria vesca, Glycine max, Gossypium hirsutum, (Gossypium arboreum, Gossypium herbaceum, Gossypium vitifolium), Helianthus annuus, Hevea brasiliensis, Hordeum vulgare, Humulus lupulus, Ipomoea batatas, Juglans regia, Lens culinaris, Linum usitatissimum, Lycopersicon lycopersicum, Malus spec., Manihot esculenta, Medicago sativa, Musa spec., Nicotiana tabacum (N.rustica), Olea europaea, Oryza sativa, Phaseolus lunatus, Phaseolus vul-

garis, Picea abies, Pinus spec., Pisum sativum, Prunus avium, Prunus persica, Pyrus communis, Ribes sylvestre, Ricinus communis, Saccharum officinarum, Secale cereale, Solanum tuberosum, Sorghum bicolor (s. vulgare), Theobroma cacao, Trifolium pratense, Triticum aestivum, Triticum durum, Vicia faba, Vitis vinifera, Zea mays, or cyanobacteria which can be transformed readily, such as Synechocystis, into which the sequence encoding a polypeptide according to the invention has been incorporated by transformation. These transgenic organisms thus show increased tolerance to compounds which inhibit the polypeptide according to the invention. "Knock-out" mutants in which the analogous CLProtease gene which is naturally present in this organism has been selectively switched off may also be used.

However, the abovementioned embodiment of the method according to the invention can also be used for identifying substances with a growth-regulatory action. In this context, the transgenic organism employed is a plant. The method for identifying substances with growth-regulatory activity thus comprises the following steps:

i. generating a transgenic plant comprising a nucleic acid sequence according to the invention encoding CLP, wherein CLP is expressed recombinantly;

ii. applying a test substance to the transgenic plant of i) and to a nontransgenic plant of the same variety,

iii. determining the growth or the viability of the transgenic plant and the nontransgenic plant after application of the test compound, and

iv. selecting test substances which bring about a reduced growth of the nontransgenic plant in comparison with the growth of the transgenic plant.

Here, step iv) involves the selection of test compounds which bring about a modified growth of the nontransgenic organism in comparison with the growth of the transgenic organism. Modified growth is understood as meaning, in this context, inhibition of the vegetative growth of the plants, which can manifest itself in particular in reduced longitudinal growth. Accordingly, the treated plants show stunted growth; moreover, their leaves are darker in color. In addition, modified growth is also understood as meaning a change in the course of maturation over time, the inhibition or promotion of lateral branched growth of the plants, shortened or extended developmental stages, increased standing ability, the growth of larger amounts of buds, flowers, leaves, fruits, seed kernels, roots and tubers, an increased sugar content in plants such as sugarbeet, sugar cane and citrus fruit, an increased protein content in plants such as cereals or soybean, or stimulation of the latex flow in rubber trees. The skilled worker is familiar with the detection of such modified growth.

It is also possible, in the method according to the invention, to employ a plurality of test compounds in a method according to the invention. If a group of test compounds affect the target, then it is either possible directly to isolate the individual test compounds or to divide the group of test compounds into a variety of subgroups, for example when it consists of a multiplicity of different components, in order to thus reduce the number of the different test compounds in the method according to the invention. The method according to the invention is then repeated with the individual test compound or the relevant subgroup of test compounds. Depending on the complexity of the sample, the above-described steps can be carried out repeatedly, preferably until the subgroup identified in accordance with the method according to the invention only comprises a small number of test compounds, or indeed just one test compound.

All of the above-described methods for identifying inhibitors with herbicidal or growth-regulatory activity are hereinbelow referred to as "methods according to the invention".

All of the compounds which have been identified via the methods according to the invention can subsequently be tested in vivo for their herbicidal and growth-regulatory activity. One possibility of testing the compounds for herbicidal action is to use duckweed, Lemna minor, in microtiter plates. Parameters which can be measured are changes in the chlorophyll content and the photosynthesis rate. It is also possible to apply the compound directly to undesired plants, it being possible to identify the herbicidal action for example via restricted growth.

The method according to the invention can advantageously also be carried out in high-throughput methods, known as HTS, which makes possible the simultaneous testing of a multiplicity of different compounds.

The use of supports which contain one or more of the nucleic acid molecules according to the invention, one or more of the vectors containing the nucleic acid sequence according to the invention, one or more transgenic organisms containing at least one of the nucleic acid sequences according to the invention or one or more (poly)peptides encoded via the nucleic acid sequences according to the invention lends itself to carrying out HTS in practice.

Supports which contain one or more of the NCLP sequences, one or more of the vectors comprising the NCLP sequences one or more transgenic organisms containing at least one NCLP sequences or one or more (poly)peptides encoded by the NCLP sequences are part of the present invention.

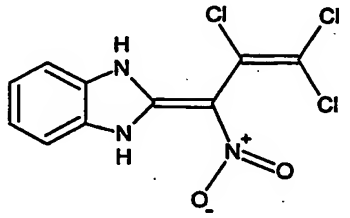
The support used can be solid or liquid, but is preferably solid and especially preferably a microtiter plate. The abovementioned supports also form part of the subject matter of the present invention. In accordance with the most widely used technique, 96-well,

384-well and 1536-well microtiter plates which, as a rule, can comprise volumes of 200 □l, are used. Besides the microtiter plates, the further components of an HTS system which match the corresponding microtiter plates, such as a large number of instruments, materials, automatic pipetting devices, robots, automated plate readers and plate washers, are commercially available.

In addition to the HTS systems based on microtiter plates, what are known as "free-format assays" or assay systems where no physical barriers exist between the samples, as described, for example, in Jayaickreme et al., Proc. Natl. Acad. Sci U.S.A. 19 (1994) 161418; Chelsky, "Strategies for Screening Combinatorial Libraries", First Annual Conference of The Society for Biomolecular Screening in Philadelphia, Pa. (Nov. 710, 1995); Salmon et al., Molecular Diversity 2 (1996), 5763 and US 5,976,813, may also be used.

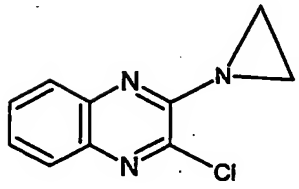
The invention furthermore relates to herbicidally active compounds identified by the methods according to the invention. These compounds are hereinbelow referred to as "selected compounds". They have a molecular weight of less than 1000 g/mol, advantageously less than 500 g/mol, preferably less than 400 g/mol, especially preferably less than 300 g/mol. Herbicidally active compounds have a Ki value of less than 1 mM, preferably less than 1 µM, especially preferably less than 0.1µM, very especially preferably less than 0.01 µM.

Examples for herbicidally active compounds identified with the above mentioned HTS methods are the compounds of the formula:



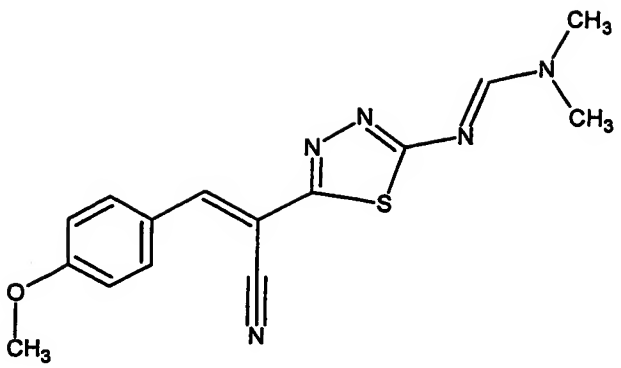
formula (I)

or



formula (II)

or



formula (III).

The invention furthermore relates to compounds with growth-regulatory activity identified by the methods according to the invention. These compounds too are hereinbelow referred to as "selected compounds".

Naturally, the selected compounds can also be present in the form of their agriculturally useful salts. Agriculturally useful salts which are suitable are mainly the salts of those cations, or the acid addition salts of those acids, whose cations, or anions, do not adversely affect the herbicidal action of the herbicidally active compounds identified via the methods according to the invention.

If the selected compounds contain asymmetrically substituted □-carbon atoms, they may furthermore also be present in the form of racemates, enantiomer mixtures, pure enantiomers or, if they have chiral substituents, also in the form of diastereomer mixtures.

The selected compounds can be chemically synthesized substances or substances produced by microbes and can be found, for example, in cell extracts of, for example, plants, animals or microorganisms. The reaction mixture can be a cell-free extract or comprise a cell or cell culture. Suitable methods are known to the skilled worker and are described generally for example in Alberts, Molecular Biology the cell, 3rd Edition (1994), for example chapter 17. The selected compounds may also originate from comprehensive substance libraries.

Candidate test compounds can be expression libraries such as, for example, cDNA expression libraries, peptides, proteins, nucleic acids, antibodies, small organic substances, hormones, PNAs or the like (Milner, Nature Medicin 1 (1995), 879–880; Hupp, Cell. 83 (1995), 237–245; Gibbs, Cell. 79 (1994), 193–198 and references cited therein).

The selected compounds can be used for controlling undesired vegetation and/or as growth regulators. Herbicidal compositions comprising the selected compounds afford very good control of vegetation on noncrop areas. In crops such as wheat, rice, maize,

soybean and cotton, they act against broad-leaved weeds and grass weeds without inflicting any significant damage on the crop plants. This effect is observed in particular at low application rates. The selected compounds can be used for controlling the harmful plants which have already been mentioned above.

Depending on the application method in question, selected compounds, or herbicidal compositions comprising them, can advantageously also be employed in a further number of crop plants for eliminating undesired plants. Examples of suitable crops are:

Allium cepa, Ananas comosus, Arachis hypogaea, Asparagus officinalis, Beta vulgaris spec. altissima, Beta vulgaris spec. rapa, Brassica napus var. napus, Brassica napus var. napobrassica, Brassica rapa var. silvestris, Camellia sinensis, Carthamus tinctorius, Carya illinoinensis, Citrus limon, Citrus sinensis, Coffea arabica (Coffea canephora, Coffea liberica), Cucumis sativus, Cynodon dactylon, Daucus carota, Elaeis guineensis, Fragaria vesca, Glycine max, Gossypium hirsutum, (Gossypium arboreum, Gossypium herbaceum, Gossypium vitifolium), Helianthus annuus, Hevea brasiliensis, Hordeum vulgare, Humulus lupulus, Ipomoea batatas, Juglans regia, Lens culinaris, Linum usitatissimum, Lycopersicon lycopersicum, Malus spec., Manihot esculenta, Medicago sativa, Musa spec., Nicotiana tabacum (N.rustica), Olea europaea, Oryza sativa, Phaseolus lunatus, Phaseolus vulgaris, Picea abies, Pinus spec., Pisum sativum, Prunus avium, Prunus persica, Pyrus communis, Ribes sylvestre, Ricinus communis, Saccharum officinarum, Secale cereale, Solanum tuberosum, Sorghum bicolor (s. vulgare), Theobroma cacao, Trifolium pratense, Triticum aestivum, Triticum durum, Vicia faba, Vitis vinifera, Zea mays.

In addition, the selected compounds can also be used in crops which tolerate the action of herbicides owing to breeding, including recombinant methods. The generation of such crops is described hereinbelow.

The invention furthermore relates to a method of preparing the herbicidal or growth-regulatory composition which has already been mentioned above, which comprises formulating selected compounds with suitable auxiliaries to give crop protection products.

The selected compounds can be formulated for example in the form of directly sprayable aqueous solutions, powders, suspensions, also highly concentrated aqueous, oily or other suspensions or suspoemulsions or dispersions, emulsifiable concentrates, emulsions, oil dispersions, pastes, dusts, materials for spreading or granules, and applied by means of spraying, atomizing, dusting, spreading or pouring. The use forms depend on the intended use and the nature of the selected compounds; in any case, they should guarantee the finest possible distribution of the selected compounds. The herbicidal compositions comprise a herbicidally active amount of at least one selected

compound and auxiliaries conventionally used in the formulation of herbicidal compositions.

For the preparation of emulsions, pastes or aqueous or oily formulations and dispersible concentrates (DC), the selected compounds can be dissolved or dispersed in an oil or solvent, it being possible to add further formulation auxiliaries for homogenization. However, it is also possible to prepare liquid or solid concentrates from selected compound, if appropriate solvents or oil and, optionally, further auxiliaries comprising liquid or solid concentrates, and these concentrates are suitable for dilution with water. The following can be mentioned: emulsifiable concentrates (EC, EW), suspensions (SC), soluble concentrates (SL), dispersible concentrates (DC), pastes, pills, wettable powders or granules, it being possible for the solid formulations either to be soluble or dispersible (wettable) in water. In addition, suitable powders or granules or tablets can additionally be provided with a solid coating which prevents abrasion or premature release of the active ingredient.

In principle, the term "auxiliaries" is understood as meaning the following classes of compounds: antifoams, thickeners, wetting agents, tackifiers, dispersants, emulsifiers, bactericides and/or thixotropic agents. The skilled worker is familiar with the meaning of the abovementioned agents.

SLs, EWs and ECs can be prepared by simply mixing the ingredients in question; powders can be prepared by mixing or grinding in specific types of mills (for example hammer mills). DCs, SCs and SEs are usually prepared by wet milling, it being possible to prepare an SE from an SC by addition of an organic phase which may comprise further auxiliaries or selected compounds. The preparation is known. Powders, materials for spreading and dusts can advantageously be prepared by mixing or cogrinding the active substances together with a solid carrier. Granules, for example coated granules, impregnated granules and homogeneous granules, can be prepared by binding the selected compounds to solid carriers. The skilled worker is familiar with further details regarding their preparation, which are mentioned for example in the following publications: US 3,060,084, EP-A 707445 (for liquid concentrates), Browning, "Agglomeration", Chemical Engineering, Dec. 4, 1967, 147-48, Perry's Chemical Engineer's Handbook, 4th Ed., McGraw-Hill, New York, 1963, pages 8-57 and et seq. WO 91/13546, US 4,172,714, US 4,144,050, US 3,920,442, US 5,180,587, US 5,232,701, US 5,208,030, GB 2,095,558, US 3,299,566, Klingman, Weed Control as a Science, John Wiley and Sons, Inc., New York, 1961, Hance et al., Weed Control Handbook, 8th Ed., Blackwell Scientific Publications, Oxford, 1989 and Mollet, H., Grubemann, A., Formulation technology, Wiley VCH Verlag GmbH, Weinheim (Federal Republic of Germany), 2001.

The skilled worker is familiar with a multiplicity of inert liquid and/or solid carriers which are suitable for the formulations according to the invention, such as, for example, liquid additives such as mineral oil fractions of medium to high boiling point such as kerosene or diesel oil, furthermore coal tar oils and oils of vegetable or animal origin, aliphatic, cyclic and aromatic hydrocarbons, for example paraffin, tetrahydrophthalene, alkylated naphthalenes or their derivatives, alkylated benzenes or their derivatives, alcohols such as methanol, ethanol, propanol, butanol and cyclohexanol, ketones such as cyclohexanone, or strongly polar solvents, for example amines such as N-methylpyrrolidone or water.

Examples of solid carriers are mineral earths such as silicas, silica gels, silicates, talc, kaolin, limestone, lime, chalk, bole, loess, clay, dolomite, diatomaceous earth, calcium sulfate, magnesium sulfate, magnesium oxide, ground synthetic materials, fertilizers such as ammonium sulfate, ammonium phosphate, ammonium nitrate, ureas and products of vegetable origin such as cereal meal, tree bark meal, wood meal and nutshell meal, cellulose powders or other solid carriers.

The skilled worker is familiar with the multiplicity of surface-active substances (surfactants) which are suitable for the formulations according to the invention such as, for example, alkali metal salts, alkaline earth metal salts or ammonium salts of aromatic sulfonic acids for example lignosulfonic acid, phenolsulfonic acid, naphthalenesulfonic acid, and dibutylnaphthalenesulfonic acid, and of fatty acids, of alkyl- and alkylarylsulfonates, of alkyl sulfates, lauryl ether sulfates and fatty alcohol sulfates, and salts of sulfated hexa-, hepta- and octadecanols and of fatty alcohol glycol ethers, condensates of sulfonated naphthalene and its derivatives with formaldehyde, condensates of naphthalene or of the naphthalenesulfonic acids with phenol and formaldehyde, polyoxyethylene octylphenol ether, ethoxylated isooctyl-, octyl- or nonylphenol, alkylphenyl polyglycol ethers, tributylphenyl polyglycol ether, alkylaryl polyether alcohols, isotridecyl alcohol, fatty alcohol/ethylene oxide condensates, ethoxylated caster oil, polyoxyethylene alkyl ethers or polyoxypropylene alkyl ethers, lauryl alcohol polyglycol ether acetate, sorbitol esters, lignosulfite waste liquors or methylcellulose.

The herbicidal compositions, or the selected compounds, can be applied pre- or post-emergence. If the selected compounds are less well tolerated by certain crop plants, application techniques may be used in which the selected compounds are sprayed, with the aid of the spraying apparatus, in such a way that they come into as little contact, if any, with the leaves of the sensitive crop plants while the selected compounds reach the leaves of undesired plants which grow underneath, or the bare soil surface (post-directed, lay-by).

Depending on the intended purpose of the control measures, the season, the target plants and the growth stage, the application rates of selected compounds amount to 0.001 to 3.0, preferably 0.01 to 1.0 kg/ha.

The invention is illustrated in greater detail by the examples which follow, which are not to be considered as limiting.

General DNA manipulation and cloning methods

Cloning methods such as, for example, restriction cleavages, agarose gel electrophoreses, purification of DNA fragments, transfer of nucleic acids to nitrocellulose and nylon membranes, linking DNA fragments, transformation of Escherichia coli cells, growing bacterium and sequence analyses of recombinant DNA were carried out as described by Sambrook et al. (1989) (Cold Spring Harbor Laboratory Press: ISBN 0-87969-309-6) and Ausubel, F.M. et al., Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley-Interscience (1994); ISBN 0-87969-309-6.

Molecular-biological standard methods for plants and plant transformation methods are described in Schultz et al., Plant Molecular Biology Manual, Kluwer Academic Publishers (1998), Reither et al., Methods in Arabidopsis Research, World scientific press (1992) and Arabidopsis: A Laboratory Manual (2001), ISBN 0-87969-573-0.

The bacterial strains used hereinbelow (E. coli DH5, XL-1 blue) were obtained from Stratagene, BRL Gibco or Invitrogen, Carlsberg, CA. The vectors used for cloning were pUC 19 from Amersham Pharmacia (Freiburg) and the vector pBinAR (Höfgen and Willmitzer, Plant Science 66, 1990, 221-230).

Example 1: Generation of a cDNA library in the plant transformation vector

To generate a cDNA library (hereinbelow termed "binary cDNA library") in a vector which can be used directly for transforming plants, mRNA was isolated from a variety of plant tissues and transcribed into double-stranded cDNA using the cDNA Synthese Kit (Amersham Pharmacia Biotech, Freiburg). The cDNA first-strand synthesis was carried out using T12-18 oligonucleotides following the manufacturer's instructions. After size fractionation and the ligation of EcoRI-NotI adapters following the manufacturer's instructions and filling up the overhangs with Pfu DNA polymerase (Stratagene), the cDNA population was normalized. The method of Kohci et al, 1995, Plant Journal 8, 771-776 was followed, the cDNA being amplified by PCR with the oligonucleotide N1 under the conditions given in Table 1.

Table 1

Temperature [°C]	Time [sec]	Number of cycles
94	300	1
94	8	10
52	60	
72	180	
94	8	10
50	60	
72	180	
94	8	10
48	60	
72	180	
72	420	1

The resulting PCR product was bound to the column matrix of the PCR purification kit (Qiagen, Hilden) and eluted with 300 mM NaP buffer, pH 7.0, 0.5 mM EDTA, 0.04% SDS. The DNA was denatured for 5 minutes in a boiling water bath and subsequently renatured for 24 hours at 60oC. 50µl of the DNA were applied to a hydroxylapatite column and the column was washed 3 times with 1 ml of 10 mM NaP buffer, pH 6.8. The bound single-stranded DNA was eluted with 130 mM NaP buffer, pH 6.8, precipitated with ethanol and dissolved in 40 µl of water. 20 µl of growth were used for a further PCR amplification as described above. After further ssDNA concentration, a third PCR amplification was carried out as described above.

The plant transformation vector for taking up the cDNA population which had been generated as described above was generated via restriction enzyme cleavage of the vector pUC18 with SbfI and BamHI, purification of the vector fragment followed by filling up the overhangs with Pfu DNA polymerase and relegation with T4 DNA ligase (Stratagene). The resulting construct is hereinbelow termed pUC18SbfI-.

The vector pBinAR was first cleaved with NotI, the ends were filled up and the vector was relegated, cleaved with SbfI, the ends were filled up and the vector was relegated and subsequently cleaved with EcoRI and HindIII. The resulting fragment was ligated into a derivative of the binary plant transformation vector pPZP (Hajdukiewicz,P, Svab, Z, Maliga, P., (1994) Plant Mol Biol 25:989-994) which makes possible the transformation of plants by means of agrobacterium and mediates kanamycin resistance in transgenic plants. The construct generated thus is hereinbelow termed pSun12/35S.

pUC18SbfI- was used as template in a polymerase chain reaction (PCR) with the oligonucleotides V1 and V2 (see Table 2) and Pfu DNA polymerase. The resulting fragment was ligated into the SmaI-cut pSun12/35S, giving rise to pSunblues2. Following cleavage with NotI, dephosphorylation with shrimp alkaline phosphatase (Roche Diagnostics, Mannheim) and purification of the vector fragment, pSunblues2 was ligated with the normalized, likewise NotI-cut cDNA population. Following transformation into E.coli XI-1blue (Stratagene), the resulting clones were deposited into microtiter plates. The binary cDNA library contains cDNAs in "sense"- and in "antisense" orientation under the control of the cauliflower mosaic virus 35S promoter, and, after transformation into tobacco plants, these cDNAs can, accordingly, lead to "cosuppression" and "antisense" effects.

Table 2: Oligonucleotides used

Oligonucleotide	Nucleic acid sequence
N1	5'-AGAATTCGCGGCCGCT-3' (SEQ ID NO:23)
V1 (PWL93not)	5'-CTCATGCGGCCGCGCGCAACGCAATTAATGTG-3' (SEQ ID NO:24)
V2 (pWL92)	5'-TCATGCGGCCGCGAGATCCAGTTCGATGTAAC-3' (SEQ ID NO:25)
G1 (35S)	5'-GTGGATTGATGTGATATCTCC-3' (SEQ ID NO:26)
G2 (OCS)	5'-GTAAGGATCTGAGCTACACAT-3' (SEQ ID NO:27)

Example 2: Transformation and analysis of tobacco plants

Selected clones of the binary cDNA library were transformed into Agrobacterium tumefaciens C58C1:pGV2260 and (Deblaere et al., Nucl. Acids. Res. 13(1984), 4777-4788) and incubated with Streptomycin/Spectinomycin selection. The material used for the transformation of tobacco plants (Nicotiana tabacum cv. Samsun NN) with one of the binary clones as depicted in table 3 was an overnight culture of a positively transformed agrobacterial colony diluted with YEB medium to OD600 = 0.8-1.6. Leaf discs of sterile plants (approx. 1 cm2 each) were incubated for 5-10 minutes with a 1:50 agrobacterial dilution in a Petri dish. This was followed by incubation in the dark for 2 days at 25°C on Murashige-Skoog medium (Physiol. Plant. 15(1962), 473) supplemented with 2% sucrose (2MS medium) and 0.8% Bacto agar. The cultivation was continued after 2 days at a 16-hour-light/8-hour-darkness photoperiod and continued in a weekly rhythm on MS medium supplemented with 500mg/l Claforan (cefotaxime sodium), 50mg/l kanamycin, 1mg/l benzylaminopurin (BAP), 0.2mg/l naphthylacetic acid and 1.6g/l glucose. Regenerated shoots were transferred onto an MS medium supplemented with kanamycin and Claforan. Transgenic plants of lines as depicted in table 3 were generated in this manner.

Table 3: Plant lines generated

Partial cDNA with phenotype in transgenic tobacco	Plant line	Corresponding full length cDNA	Function
SEQ ID NO: 20	E_0000013511	SEQ ID NO: 3	ClpP2 protease
SEQ ID NO:19	E_0000008893	SEQ ID NO:11	ClpP5=ClpR1 protease
SEQ ID NO:21	E_0000012393	-	ClpP6 protease
-	-	SEQ ID NO:17	ClpP-like protease

The integration of the clone cDNA into the genome of the transgenic lines was detected via PCR with the oligonucleotides G1 and G2 (see Table 2) and genomic DNA prepared from the transgenic lines in question. To this end, TAKARA Taq DNA polymerase was preferably employed for this purpose, following the manufacturer's instructions (MoBiTec, Göttingen). The cDNA clone of the binary cDNA library, which clone had been used for the transformation, acted as template for a PCR reaction as the positive control. PCR products with an identical size or, if appropriate, identical cleavage patterns which were obtained after cleavage with a variety of restriction enzymes acted as proof that the corresponding cDNA had been integrated. In this manner, the insert of clones were detected in the respective transgenic plant lines (as depicted in table 3) with the belowmentioned phenotypes.

After the shoots had been transferred into soil, the plants were observed for 2-20 weeks in the greenhouse for the manifestation of phenotypes. It emerged that transgenic plants of lines E_0000012393, E_0000013511 and E_0000008893 were similar in phenotype. The plants showed severe chlorosis and concomitant growth retardation with respect to wild type plants after 2 weeks.

Example 3: Sequence analysis of the clones

SEQ ID NO:19 was fully sequenced and used for the detection of the corresponding full length clone SEQ ID NO:11. SEQ ID NO:11 is identical to ~~nt002050071r~~ SEQ ID NO:19 in the overlapping region. An open reading frame of 867 nt (pos. 2-1162) encodes for 387 amino acids (SEQ ID NO: 4) with highest identity to ClpR1 from Arabidopsis thaliana. Sequence homology suggests that the 5'-ends of SEQ ID NO:11 and ClpR1 from Arabidopsis thaliana are very diverse and that nt006066004r is close to being full size with respect to ClpR1. MS-Analysis of isolated ClpP Proteins from Arabidopsis indicate, that the mature ClpR1 is several kDal shorter as suggested by the cDNA Sequence (Peltier et al. 2001, The Journal of Biological Chemistry 276, 99.16318-16327).

SEQ ID NO:20 was fully sequenced and used for the detection of the corresponding full length clone, SEQ ID NO: 3. SEQ ID NO:3 is identical to SEQ ID NO:20 in the overlapping region. An open reading frame of 867 nt (pos. 11-877) encodes for 289 amino acids (SEQ ID NO:4) with highest identity to ClpP2 from Arabidopsis thaliana.

SEQ ID NO:21 was fully sequenced . The partial cDNA Sequence of 602 nt contains an open reading frame of 186 nt (nt 8-193) encoding for 62 amico acids (SEQ ID NO:22). This partial polypeptide shows highest identity to ClpP6 from Arabidopsis thaliana (SEQ ID NO:9)

A further ClpP-homolog cDNA of 906 nt (SEQ ID NO:17) was identified. An open reading frame of 711 nt (pos. 45-755) encodes for 237 amino acids (SEQ ID NO:18) with highest identity to a ClpP-like protein from Arabidopsis thaliana (GeneBank Acc. No. AK118523).

Thus, it was shown for the first time and in a surprising manner that the natural expression of nuclear encoded Clp protease encoding genes is essential for plants and that reduced expression leads to damage as depicted by the phenotypes mentioned in Example 2 demonstrating the suitability of nuclear encoded Clp-proteases as target for herbicides.

Example 4: Expression in E.coli

In order to generate active protein with nuclear encoded Clp-protease activity fragments of SEQ ID NO:11, SEQ ID NO:3 and of SEQ ID NO:9 were subcloned into the expression vector pQE60 (Quiagen, Hilden, Germany). To this end the oligonucleotides displayed in tab. 4 where used to amplify via polymerase chain reaction cDNA fragments that contain NcoI and BglII restriction sites. The PCR was carried out in 36 cycles following standard conditions (for example as described by Sambrook, J. et al. (1989) "Molecular cloning: A laboratory manual", Cold Spring Harbor Laboratory Press), the annealing temperatures being between 45 and 55°C and the polymerization time being in each case 60 seconds per 1000bp. Cutting the cDNA fragments with NcoI and BglII restriction enzymes and ligation into pQE60 cut with the same enzymes delivered expression plasmids that were transformed into E. coli. Expression was performed in E. coli TOP 10F strains (Invitrogen, Karlsruhe, Germany) following induction with IPTG. Standard protocols (Invitrogen) were followed.

Enzyme preperations were achieved by breaking cells in a French-Press in 100mM Tris/HCl, pH 7.4, 2.5 mM EDTA, 1% Triton X-100.

The expression products were purified by affinity chromatography on Ni-agarose where appropriate. The manufacturer's instructions were followed (Qiagen).

Table 4

Construct	Primer (Nucleic acid sequence)
NT_ClpR1[1)]	5'-TATACCATGGATTTGCCATCTTTG-3' (SEQ ID NO:28) 5'-ATAGATCTCACCTGGAGCCAG-3' (SEQ ID NO:29)
Nt_ClpP2[1)]	5'-GAGCCCATGGCAAGAGGAG -3' (SEQ ID NO:30) 5'-ATAGATCTTTCTAGCTTGAACC-3' (SEQ ID NO:31)
AT_ClpP6[2)]	5'-TCAGCCATGGCCCCTGGAGGAC -3'(SEQ ID NO:32) 5'-TAAGATCTTCAGTATTCTGTTTCC-3' (SEQ ID NO:33)

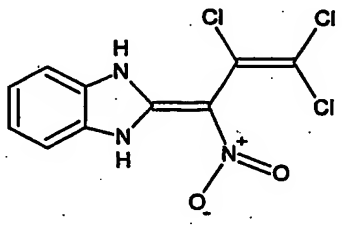
1) Template: Nicotiana tabacum cDNA library
2) Template: Arabidopsis thaliana cDNA library

Example 5: Activity assay

Isolated ClpP activity can be measured as described (Woo et al. 1989 The Journal of Biological Chemistry 264, pp.2088-2091) by using fluoregenic substrates such as N-Suc-Leu-Tyr-(7-amino-4-methylcoumarine) (SLT-AMC). The proteolytic cleavage deliberates 7-amino-4-methylcoumarin, which can be detected fluorimetrically (emission at 460nm by exitation at 390 nm).
Standard assays contain: 50mM Tris/HCl, pH 8.0, 25mM $MgCl_2$, 1mM SLT-AMC and 1-100 µg ClpP Enzyme.

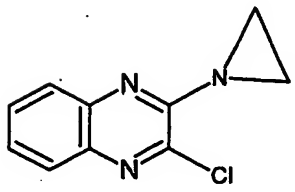
The assay is suitable in for high throughput screening in 96well and 384 well format.

Screening according to the above mentioned assay provided the following compounds of the formula:



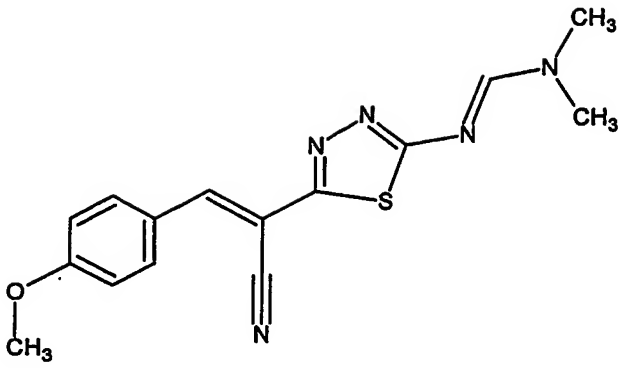
formula (I)

or



formula (II)

or



formula (III).

showing a inhibition of the enzyme of:

compound of formula	IC50
(I)	2.3E-05
(II)	1.9E-05
(III)	2.5E-05

Sequence Listing

Sequence	Function	Organism
SEQ ID NO:1 (nucleic acid sequence)	ClpP1	Arabidopsis thanliana
SEQ ID NO:2 (amino acid sequence)	ClpP1	Arabidopsis thanliana
SEQ ID NO:3 (nucleic acid sequence)	ClpP2	Nicotiana tabacuum
SEQ ID NO:4 (amino acid sequence)	ClpP2	Nicotiana tabacuum
SEQ ID NO:5 (nucleic acid sequence)	ClpP3	Arabidopsis thanliana
SEQ ID NO:6 (amino acid sequence)	ClpP3	Arabidopsis thanliana
SEQ ID NO:7 (nucleic acid sequence)	ClpP4	Arabidopsis thanliana
SEQ ID NO:8 (amino acid sequence)	ClpP4	Arabidopsis thanliana
SEQ ID NO:9 (nucleic acid sequence)	ClpP6	Arabidopsis thanliana

SEQ ID NO:10 (amino acid sequence)	ClpP6	Arabidopsis thanliana
SEQ ID NO:11 (nucleic acid sequence)	ClpR1	Nicotiana tabacuum
SEQ ID NO:12 (amino acid sequence)	ClpR1	Nicotiana tabacuum
SEQ ID NO:13 (nucleic acid sequence)	ClpR3	Arabidopsis thanliana
SEQ ID NO:14 (amino acid sequence)	ClpR3	Arabidopsis thanliana
SEQ ID NO:15 (nucleic acid sequence)	ClpR4	Arabidopsis thanliana
SEQ ID NO:16 (amino acid sequence)	ClpR4	Arabidopsis thanliana
SEQ ID NO:17 (nucleic acid sequence)	ClpP like	Arabidopsis thanliana
SEQ ID NO:18 (amino acid sequence)	ClpP like	Arabidopsis thanliana
SEQ ID NO:19 (nucleic acid sequence) (fragment)	ClpR1	Nicotiana tabacuum
SEQ ID NO:20 (nucleic acid sequence) (fragment)	ClpP2	Nicotiana tabacuum
SEQ ID NO:21 (nucleic acid sequence) (fragment)	ClpP6	Nicotiana tabacuum
SEQ ID NO:22 (amino acid sequence) (fragment)	ClpP6	Nicotiana tabacuum

SEQ ID NO:23-33: Primer (nucleic acid sequences)

We claim:

1. The use of nuclear encoded Clp-protease in a method for identifying herbicides.

2. The use as claimed in claim 1, wherein the Clp-protease is

a) selected from the group consisting of ClpP1-protease, ClpP2-protease, ClpP3-protease, ClpP4-protease and ClpP6-protease; or

b) selected from the group consisting of ClpR1-protease, ClpR3-protease, ClpR4-protease; or

c) ClpP-like-protease.

3. A plant nucleic acid sequence encoding a ClpP2-protease comprising:

a) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:3, or

b) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:4 by backtranslating, or

c) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:3 which has an identity with SEQ ID NO:3 of has at least 66%.

4. A plant nucleic acid sequence encoding a ClpR1-protease comprising:

a) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:11, or

b) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:12 by backtranslating, or

c) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:11 which has an identity with SEQ ID NO:11 of has at least 69%.

5. A plant nucleic acid sequence encoding a ClpP-like-protease comprising:

a) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:17, or

b) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:18 by backtranslating, or

c) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:17 which has an identity with SEQ ID NO:17 of has at least 67%.

6. A polypeptide with the activity of a nuclear encoded Clp-protease, encoded by a nucleic acid molecule as claimed in claim 3, 4 or 5.

7. An expression cassette comprising

a) genetic control sequences in operable linkage with a nucleic acid sequence as claimed in claim 3, 4 or 5; or

b) additional functional elements, or

c) a combination of a) and b).

8. A vector comprising an expression cassette as claimed in claim 7.

9. A transgenic organism comprising at least one nucleic acid sequence as claimed in claim 4, 5 or 6 encoding a polypeptide with the activity of a Clp-protease, an expression cassette as claimed in claim 7 or a vector as claimed in claim 8, selected from among bacteria, yeasts, fungi, animal cells or plant cells.

10. A method for identifying substances with herbicidal activity, comprising the following steps:

i. bringing a nuclear encoded Clp-protease into contact with one or more test compounds under conditions which permit the test compound(s) to bind to the nucleic acid molecule encoding Clp-protease or to the nuclear encoded Clp-protease, and

ii. detecting whether the test compound binds to the Clp-protease of i), or

iii. detecting whether the test compound reduces or blocks the enzymatic or biological activity of the Clp-protease of i), or

iv. detecting whether the test compound reduces or blocks the transcription, translation or expression of the Clp-protease of i).

11. A method as claimed in claim 10, wherein the Clp-protease is

a) selected from the group consisting of ClpP1-protease, ClpP2-protease, ClpP3-protease, ClpP4-protease and ClpP6-protease; or

b) selected from the group consisting of ClpR1-protease, ClpR3-protease, ClpR4-protease; or

c) ClpP-like-protease.

12. A method as claimed in claim 10, wherein

a) the ClpP1-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:1, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:2 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:1 which has an identity with SEQ ID NO:1 of has at least 50%;

b) the ClpP2-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:3, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:4 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:3 which has an identity with SEQ ID NO:3 of has at least 50%;

c) the ClpP3-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:5, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:6 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:5 which has an identity with SEQ ID NO:5 of has at least 50%;

d) the ClpP4-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:7, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:8 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:7 which has an identity with SEQ ID NO:7 of has at least 50%;

e) the ClpP6-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:9, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:10 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:9 which has an identity with SEQ ID NO:9 of has at least 50%;

f) the ClpR1-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:11, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:12 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:11 which has an identity with SEQ ID NO:11 of has at least 50%;

g) the ClpR3-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:13, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:14 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:13 which has an identity with SEQ ID NO:13 of has at least 50%;

h) the ClpR4-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:15, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:16 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:15 which has an identity with SEQ ID NO:15 of has at least 50%;

i) the ClpP like-protease is encoded by a nucleic acid sequence which comprises:

i) a nucleic acid sequence with the nucleic acid sequence shown in SEQ ID NO:17, or

ii) a nucleic acid sequence which, owing to the degeneracy of the genetic code, can be deduced from the amino acid sequence shown in SEQ ID NO:18 by back translating, or

iii) a functional equivalent of nucleic acid sequence shown in SEQ ID NO:17 which has an identity with SEQ ID NO:17 of has at least 50%;

13. A method as claimed in claim 10, 11 or 12, wherein a test compound is selected which reduces or blocks the enzymatic or biological activity of Clp-protease.

14. A method as claimed in any of claims 10, 11, 12 or 13, wherein

i. either Clp-protease is expressed in a transgenic organism or an organism which naturally contains Clp-protease is grown,

ii. the Clp-protease of step i) is brought into contact with a test compound in the cell digest of the transgenic or nontransgenic organism, in partially purified form or in homogeneously purified form, and

iii. selecting a test compound which reduces or blocks the enzymatic activity of the Clp-protease of step a).

15. A method as claimed in any of claims 10, 11, 12 or 13, which comprises the following steps:

i. generating a transgenic organism comprising a nucleic acid sequence encoding Clp-protease, wherein Clp-protease is expressed recombinantly;

ii. applying a test substance to the transgenic organism of i) and to a non-transgenic organism of the same genotype,

iii. determining the growth or the viability of the transgenic plant and the non-transgenic plant after application of the test compound, and

iv. selecting test substances which bring about a reduced growth of the non-transgenic plant in comparison with the growth of the transgenic plant.

16. A method as claimed in claim 15, which is carried out in a plant organism, a cyanobacterium or proteobacterium.

17. A method for identifying substances with growth-regulatory activity, which comprises the following steps:

i. generating a transgenic plant comprising a nucleic acid sequence Clp-protease, wherein Clp-protease is expressed recombinantly;

ii. applying a test substance to the transgenic plant of i) and to a non-transgenic plant of the same variety,

iii. determining the growth or the viability of the transgenic plant and the non-transgenic plant after application of the test compound, and

iv. selecting test substances which bring about a reduced growth of the non-transgenic plant in comparison with the growth of the transgenic plant.

18. A method as claimed in any of claims 10 to 17, wherein the substances are identified in high-throughput screening method.

19. A support comprising one or more of the nucleic acid molecules as claimed in claim 3, 4, or 5 one or more expression cassettes as claimed in claim 7, one or more vectors as claimed in claim 8, one or more organisms as claimed in claim 9 or one or more (poly)peptides as claimed in claim 6.

20. A method as claimed in any of claims 10 to 18, wherein the substances are identified in High-Throughput Screening using a support as claimed in claim 19.

21. The use of a compound with herbicidal activity, identified by one of the methods as claimed in any of claims 10 to 16, 18 and 20 for controlling undesired vegetation and/or for regulating the growth of plants.

22. The use of a compound with growth-regulatory activity, identified by the method as claimed in any of claims 17, 18 or 20 for controlling undesired vegetation and/or for regulating the growth of plants.

23. A method for the preparation of an agrochemical composition, which comprises

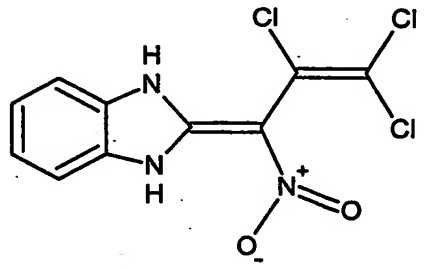
a) identifying a compound with herbicidal activity by one of the methods as claimed in any of claims 10 to 16, 18 and 20 or a compound with growth-regulatory activity as claimed in any of claims 17, 18 or 20, and

b) formulating this compound together with suitable auxiliaries to give crop protection products with herbicidal or growth-regulatory activity.

24. The use of at least one Clp-protease inhibitor identified by one of the methods as claimed in any of claims 10 to 16, 18 and 20 in a method for controlling undesired vegetation and/or for regulating the growth of plants.

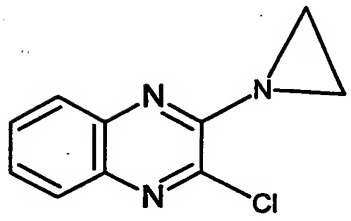
25. A method for controlling undesired vegetation and/or for regulating the growth of plants comprising treating said weeds with a herbicide, wherein said herbicide is a compound which is a inhibitor of a Clp-protease.

26. Clp-protease inhibitor of the formula:



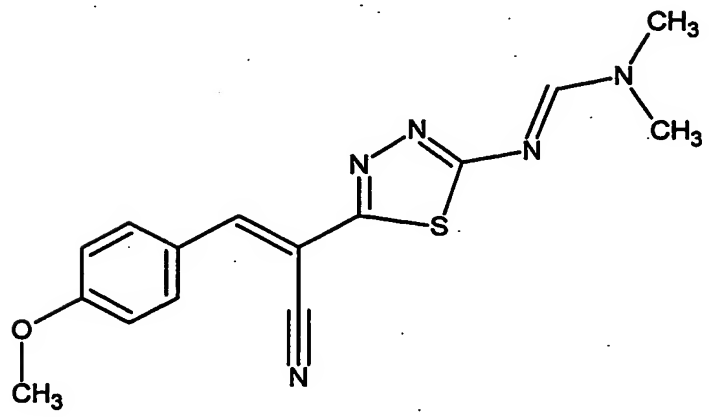
formula (I)

or



formula (II)

or



formula (III).

SEQUENCE LISTING

<110> BASF Aktiengesellschaft

<120> Clp-protease as target for herbicides

<130> 20030949

<160> 33

<170> PatentIn version 3.1

<210> 1

<211> 591

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (1)..(591)

<223>

<400> 1

```
atg cct att ggc gtt cca aaa gta cct ttt cga agt cct gga gaa gga       48
Met Pro Ile Gly Val Pro Lys Val Pro Phe Arg Ser Pro Gly Glu Gly
1               5                   10                  15

gat aca tct tgg gtt gac ata tac aac cga ctt tat cga gaa aga tta       96
Asp Thr Ser Trp Val Asp Ile Tyr Asn Arg Leu Tyr Arg Glu Arg Leu
            20                  25                  30

ttt ttt tta ggc caa gag gtt gat acc gaa atc tcg aat caa ctt att      144
Phe Phe Leu Gly Gln Glu Val Asp Thr Glu Ile Ser Asn Gln Leu Ile
        35                  40                  45

agt ctt atg ata tat ctc agt ata gaa aag gat acc aaa gat ctt tat      192
Ser Leu Met Ile Tyr Leu Ser Ile Glu Lys Asp Thr Lys Asp Leu Tyr
    50                  55                  60

ttg ttt ata aac tct cct ggt gga tgg gta ata tct gga atg gct att      240
Leu Phe Ile Asn Ser Pro Gly Gly Trp Val Ile Ser Gly Met Ala Ile
65                  70                  75                  80

tat gat act atg caa ttt gtg cga ccc gat gta cag aca ata tgc atg      288
```

Tyr Asp Thr Met Gln Phe Val Arg Pro Asp Val Gln Thr Ile Cys Met
85 90 95

gga ttg gcc gct tca ata gca tcc ttt atc cta gtc gga gga gca att 336
Gly Leu Ala Ala Ser Ile Ala Ser Phe Ile Leu Val Gly Gly Ala Ile
100 105 110

acc aaa cgt ata gca ttc cct cac gct agg gta atg atc cat caa ccc 384
Thr Lys Arg Ile Ala Phe Pro His Ala Arg Val Met Ile His Gln Pro
115 120 125

gct agt tcg ttt tat gag gca caa acg gga gaa ttt atc ttg gaa gcg 432
Ala Ser Ser Phe Tyr Glu Ala Gln Thr Gly Glu Phe Ile Leu Glu Ala
130 135 140

gaa gaa tta ctt aaa ctt cgc gaa acc atc aca agg gtt tat gta caa 480
Glu Glu Leu Leu Lys Leu Arg Glu Thr Ile Thr Arg Val Tyr Val Gln
145 150 155 160

aga acg ggc aaa cct ata tgg gtt ata tcc gaa gac atg gaa cgg gat 528
Arg Thr Gly Lys Pro Ile Trp Val Ile Ser Glu Asp Met Glu Arg Asp
165 170 175

gtt ttt atg tca gca aca gaa gcc caa gct cat gga att gtt gat ctt 576
Val Phe Met Ser Ala Thr Glu Ala Gln Ala His Gly Ile Val Asp Leu
180 185 190

gta gcg gtt caa taa 591
Val Ala Val Gln
195

<210> 2

<211> 196

<212> PRT

<213> Arabidopsis thaliana

<400> 2

Met Pro Ile Gly Val Pro Lys Val Pro Phe Arg Ser Pro Gly Glu Gly
1 5 10 15

Asp Thr Ser Trp Val Asp Ile Tyr Asn Arg Leu Tyr Arg Glu Arg Leu
20 25 30

Phe Phe Leu Gly Gln Glu Val Asp Thr Glu Ile Ser Asn Gln Leu Ile
35 40 45

Ser Leu Met Ile Tyr Leu Ser Ile Glu Lys Asp Thr Lys Asp Leu Tyr
50 55 60

Leu Phe Ile Asn Ser Pro Gly Gly Trp Val Ile Ser Gly Met Ala Ile
65 70 75 80

Tyr Asp Thr Met Gln Phe Val Arg Pro Asp Val Gln Thr Ile Cys Met
85 90 95

Gly Leu Ala Ala Ser Ile Ala Ser Phe Ile Leu Val Gly Gly Ala Ile
100 105 110

Thr Lys Arg Ile Ala Phe Pro His Ala Arg Val Met Ile His Gln Pro
115 120 125

Ala Ser Ser Phe Tyr Glu Ala Gln Thr Gly Glu Phe Ile Leu Glu Ala
130 135 140

Glu Glu Leu Leu Lys Leu Arg Glu Thr Ile Thr Arg Val Tyr Val Gln
145 150 155 160

Arg Thr Gly Lys Pro Ile Trp Val Ile Ser Glu Asp Met Glu Arg Asp
165 170 175

Val Phe Met Ser Ala Thr Glu Ala Gln Ala His Gly Ile Val Asp Leu
180 185 190

Val Ala Val Gln
195

<210> 3

<211> 1024

<212> DNA

<213> Nicotiana tabacum

<220>

<221> CDS

<222> (11)..(877)

<223>

<400> 3

gcggccgcta atg gcg gtc act ttt ccg acc acc tct tcc tcg tat cta 49
Met Ala Val Thr Phe Pro Thr Thr Ser Ser Ser Tyr Leu
1 5 10

cac tcg aga act aaa gtc cct cag cct tct tta agc tgc gcc agc aaa 97
His Ser Arg Thr Lys Val Pro Gln Pro Ser Leu Ser Cys Ala Ser Lys
15 20 25

gtt ttt gtc gga tta aga agc caa tct cct aat tct tat ggg att gca 145
Val Phe Val Gly Leu Arg Ser Gln Ser Pro Asn Ser Tyr Gly Ile Ala
30 35 40 45

```
gcg tct aat gta aat gtt gaa ttt cac aat aga gtg tac aga agt att      193
Ala Ser Asn Val Asn Val Glu Phe His Asn Arg Val Tyr Arg Ser Ile
                50                  55                  60

gaa tcc gga act aga gac agt aaa cca aca cgt gta cga gtt tcc atg      241
Glu Ser Gly Thr Arg Asp Ser Lys Pro Thr Arg Val Arg Val Ser Met
            65                  70                  75

atg ccc att ggg aca cca aga gta ccc tac aga aat cca act gag gga      289
Met Pro Ile Gly Thr Pro Arg Val Pro Tyr Arg Asn Pro Thr Glu Gly
        80                  85                  90

aca tgg cag tgg gtt gat ttg tgg aat gct ctt tac cgt gaa cgt gtt      337
Thr Trp Gln Trp Val Asp Leu Trp Asn Ala Leu Tyr Arg Glu Arg Val
    95                  100                 105

att ttc atc gga caa cac ata gat gaa gaa ttt agc aac cag ata ttg      385
Ile Phe Ile Gly Gln His Ile Asp Glu Glu Phe Ser Asn Gln Ile Leu
110                 115                 120                 125

gca aca atg ctg tat ctt gac agt att gat gat tcc aag aag ctc tac      433
Ala Thr Met Leu Tyr Leu Asp Ser Ile Asp Asp Ser Lys Lys Leu Tyr
                130                 135                 140

ctg tat atc aat ggc cct ggt ggt gat cta act cca agc atg gcc atc      481
Leu Tyr Ile Asn Gly Pro Gly Gly Asp Leu Thr Pro Ser Met Ala Ile
            145                 150                 155

tac gac aca atg caa agt ctt aaa agt gct gtt ggc acc cat tgt gtg      529
Tyr Asp Thr Met Gln Ser Leu Lys Ser Ala Val Gly Thr His Cys Val
        160                 165                 170

ggc tat gcc tac aat ctt gcc ggt ttt ctt ctt gct gct gga gaa aag      577
Gly Tyr Ala Tyr Asn Leu Ala Gly Phe Leu Leu Ala Ala Gly Glu Lys
    175                 180                 185

ggc aat cga ttt gca atg cct ctt tca agg att gca cta caa tct cca      625
Gly Asn Arg Phe Ala Met Pro Leu Ser Arg Ile Ala Leu Gln Ser Pro
190                 195                 200                 205

gct gga gct gcg cgc gga cag gct gat gat att cgc aat gaa gca gat      673
Ala Gly Ala Ala Arg Gly Gln Ala Asp Asp Ile Arg Asn Glu Ala Asp
                210                 215                 220

gaa ctt ctc aga att aga gat tac ctt ttc aag gag ttg gct gag aag      721
Glu Leu Leu Arg Ile Arg Asp Tyr Leu Phe Lys Glu Leu Ala Glu Lys
            225                 230                 235

aca ggc cag cct gtt gaa aag gtt cac aag gat tta agt cgg atg aag      769
Thr Gly Gln Pro Val Glu Lys Val His Lys Asp Leu Ser Arg Met Lys
        240                 245                 250

cga ctc aat gct caa gaa gct ctt gaa tat ggt ctt ata gac cgt ata      817
Arg Leu Asn Ala Gln Glu Ala Leu Glu Tyr Gly Leu Ile Asp Arg Ile
    255                 260                 265

gtt agg cct ccc cgt att aag gca gat gct cca cga aag gat acc aca      865
Val Arg Pro Pro Arg Ile Lys Ala Asp Ala Pro Arg Lys Asp Thr Thr
270                 275                 280                 285

gca ggt ctt ggt tagtccatac acatcgtata atttatggct gatagtggtt          917
Ala Gly Leu Gly
```

gtacgacttg cagtgttatt ttgcaatttc ttttgtttaa tctacatatt gaactctttt 977

gatctactta ttcaaaaaca tgaaatcctg agcagactag cggccgc 1024

<210> 4

<211> 289

<212> PRT

<213> Nicotiana tabacum

<400> 4

Met Ala Val Thr Phe Pro Thr Thr Ser Ser Ser Tyr Leu His Ser Arg
1 5 10 15

Thr Lys Val Pro Gln Pro Ser Leu Ser Cys Ala Ser Lys Val Phe Val
20 25 30

Gly Leu Arg Ser Gln Ser Pro Asn Ser Tyr Gly Ile Ala Ala Ser Asn
35 40 45

Val Asn Val Glu Phe His Asn Arg Val Tyr Arg Ser Ile Glu Ser Gly
50 55 60

Thr Arg Asp Ser Lys Pro Thr Arg Val Arg Val Ser Met Met Pro Ile
65 70 75 80

Gly Thr Pro Arg Val Pro Tyr Arg Asn Pro Thr Glu Gly Thr Trp Gln
85 90 95

Trp Val Asp Leu Trp Asn Ala Leu Tyr Arg Glu Arg Val Ile Phe Ile
100 105 110

Gly Gln His Ile Asp Glu Glu Phe Ser Asn Gln Ile Leu Ala Thr Met
115 120 125

Leu Tyr Leu Asp Ser Ile Asp Asp Ser Lys Lys Leu Tyr Leu Tyr Ile
130 135 140

Asn Gly Pro Gly Gly Asp Leu Thr Pro Ser Met Ala Ile Tyr Asp Thr
145 150 155 160

Met Gln Ser Leu Lys Ser Ala Val Gly Thr His Cys Val Gly Tyr Ala
165 170 175

Tyr Asn Leu Ala Gly Phe Leu Leu Ala Ala Gly Glu Lys Gly Asn Arg
180 185 190

Phe Ala Met Pro Leu Ser Arg Ile Ala Leu Gln Ser Pro Ala Gly Ala
195 200 205

Ala Arg Gly Gln Ala Asp Asp Ile Arg Asn Glu Ala Asp Glu Leu Leu
210 215 220

Arg Ile Arg Asp Tyr Leu Phe Lys Glu Leu Ala Glu Lys Thr Gly Gln
225 230 235 240

Pro Val Glu Lys Val His Lys Asp Leu Ser Arg Met Lys Arg Leu Asn
245 250 255

Ala Gln Glu Ala Leu Glu Tyr Gly Leu Ile Asp Arg Ile Val Arg Pro
260 265 270

Pro Arg Ile Lys Ala Asp Ala Pro Arg Lys Asp Thr Thr Ala Gly Leu
275 280 285

Gly

<210> 5

<211> 1124

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (2)..(931)

<223>

<400> 5
a atg gag atg agt ttg cgt ctc gct tca tct tca acc tca aac cca att 49
Met Glu Met Ser Leu Arg Leu Ala Ser Ser Ser Thr Ser Asn Pro Ile
1 5 10 15

tgt cta cta aac cct gga aaa aac ctt aat ttc cca atc cga aac cat 97
Cys Leu Leu Asn Pro Gly Lys Asn Leu Asn Phe Pro Ile Arg Asn His
20 25 30

aga atc cct aaa act tcg aaa ccc ttt tgc gtt agg tct tca atg agc 145
Arg Ile Pro Lys Thr Ser Lys Pro Phe Cys Val Arg Ser Ser Met Ser
35 40 45

ttg tct aaa cca ccc aga caa acc tta tct agt aac tgg gat gta tct 193

```
Leu Ser Lys Pro Pro Arg Gln Thr Leu Ser Ser Asn Trp Asp Val Ser
     50                  55                  60

agc ttc tcc att gat tcc gtt gct caa tct cct tca aga ctc cca agt      241
Ser Phe Ser Ile Asp Ser Val Ala Gln Ser Pro Ser Arg Leu Pro Ser
65                  70                  75                  80

ttc gaa gaa ctc gat acc acc aac atg ttg ctc cgt caa aga atc gtc      289
Phe Glu Glu Leu Asp Thr Thr Asn Met Leu Leu Arg Gln Arg Ile Val
                85                  90                  95

ttt ttg ggt tct cag gtt gat gat atg acg gcg gat ttg gtt ata agt      337
Phe Leu Gly Ser Gln Val Asp Asp Met Thr Ala Asp Leu Val Ile Ser
            100                 105                 110

cag cta ttg tta cta gat gct gag gac tca gaa aga gac att acg ctt      385
Gln Leu Leu Leu Leu Asp Ala Glu Asp Ser Glu Arg Asp Ile Thr Leu
        115                 120                 125

ttt atc aat tca ccc ggt gga tct att act gct ggg atg gga ata tat      433
Phe Ile Asn Ser Pro Gly Gly Ser Ile Thr Ala Gly Met Gly Ile Tyr
    130                 135                 140

gat gca atg aaa caa tgt aag gcg gat gta tct act gtt tgc tta ggg      481
Asp Ala Met Lys Gln Cys Lys Ala Asp Val Ser Thr Val Cys Leu Gly
145                 150                 155                 160

tta gct gca tct atg ggt gcg ttt ctt ctt gct tct ggt tca aaa ggg      529
Leu Ala Ala Ser Met Gly Ala Phe Leu Leu Ala Ser Gly Ser Lys Gly
                165                 170                 175

aaa cgg tat tgt atg cct aac tct aaa gtt atg atc cat cag cca ctt      577
Lys Arg Tyr Cys Met Pro Asn Ser Lys Val Met Ile His Gln Pro Leu
            180                 185                 190

ggt act gct gga ggc aaa gca acg gaa atg agc ata cgt ata aga gaa      625
Gly Thr Ala Gly Gly Lys Ala Thr Glu Met Ser Ile Arg Ile Arg Glu
        195                 200                 205

atg atg tac cac aag att aaa ctt aac aaa atc ttc tct aga atc act      673
Met Met Tyr His Lys Ile Lys Leu Asn Lys Ile Phe Ser Arg Ile Thr
    210                 215                 220

ggg aag cct gaa tca gag atc gaa agt gac aca gac cgt gat aac ttc      721
Gly Lys Pro Glu Ser Glu Ile Glu Ser Asp Thr Asp Arg Asp Asn Phe
225                 230                 235                 240

ttg aat cca tgg gag gcg aaa gaa tat ggt ttg atc gac gct gta atc      769
Leu Asn Pro Trp Glu Ala Lys Glu Tyr Gly Leu Ile Asp Ala Val Ile
                245                 250                 255

gat gat ggg aaa ccg gga cta atc gct cca att gga gat ggt act cct      817
Asp Asp Gly Lys Pro Gly Leu Ile Ala Pro Ile Gly Asp Gly Thr Pro
            260                 265                 270

cct cct aaa acc aaa gtc tgg gat ctt tgg aaa gtc gaa gga acc aag      865
Pro Pro Lys Thr Lys Val Trp Asp Leu Trp Lys Val Glu Gly Thr Lys
        275                 280                 285

aaa gac aac act aac ttg cca tct gag cgc tcc atg aca cag aat ggt      913
Lys Asp Asn Thr Asn Leu Pro Ser Glu Arg Ser Met Thr Gln Asn Gly
    290                 295                 300

tat gcc gcc att gaa tag aactgttgtt gcagcgttta cgccttttat            961
```

Tyr Ala Ala Ile Glu
305

atgttattct ggtggtacct gtaaccatat aacgttgcat ttcctgtgtt tgtaccattt 1021

ctctgataga ttttggaata atttgaaggc aaagatagat tattgtgtag agagctacaa 1081

atttaatgat aaattgatca tcagcactgg aaagctaaaa aaa 1124

<210> 6

<211> 309

<212> PRT

<213> Arabidopsis thaliana

<400> 6

Met Glu Met Ser Leu Arg Leu Ala Ser Ser Ser Thr Ser Asn Pro Ile
1 5 10 15

Cys Leu Leu Asn Pro Gly Lys Asn Leu Asn Phe Pro Ile Arg Asn His
20 25 30

Arg Ile Pro Lys Thr Ser Lys Pro Phe Cys Val Arg Ser Ser Met Ser
35 40 45

Leu Ser Lys Pro Pro Arg Gln Thr Leu Ser Ser Asn Trp Asp Val Ser
50 55 60

Ser Phe Ser Ile Asp Ser Val Ala Gln Ser Pro Ser Arg Leu Pro Ser
65 70 75 80

Phe Glu Glu Leu Asp Thr Thr Asn Met Leu Leu Arg Gln Arg Ile Val
85 90 95

Phe Leu Gly Ser Gln Val Asp Asp Met Thr Ala Asp Leu Val Ile Ser
100 105 110

Gln Leu Leu Leu Leu Asp Ala Glu Asp Ser Glu Arg Asp Ile Thr Leu
115 120 125

Phe Ile Asn Ser Pro Gly Gly Ser Ile Thr Ala Gly Met Gly Ile Tyr
130 135 140

Asp Ala Met Lys Gln Cys Lys Ala Asp Val Ser Thr Val Cys Leu Gly
145 150 155 160

Leu Ala Ala Ser Met Gly Ala Phe Leu Leu Ala Ser Gly Ser Lys Gly
165 170 175

Lys Arg Tyr Cys Met Pro Asn Ser Lys Val Met Ile His Gln Pro Leu
180 185 190

Gly Thr Ala Gly Gly Lys Ala Thr Glu Met Ser Ile Arg Ile Arg Glu
195 200 205

Met Met Tyr His Lys Ile Lys Leu Asn Lys Ile Phe Ser Arg Ile Thr
210 215 220

Gly Lys Pro Glu Ser Glu Ile Glu Ser Asp Thr Asp Arg Asp Asn Phe
225 230 235 240

Leu Asn Pro Trp Glu Ala Lys Glu Tyr Gly Leu Ile Asp Ala Val Ile
245 250 255

Asp Asp Gly Lys Pro Gly Leu Ile Ala Pro Ile Gly Asp Gly Thr Pro
260 265 270

Pro Pro Lys Thr Lys Val Trp Asp Leu Trp Lys Val Glu Gly Thr Lys
275 280 285

Lys Asp Asn Thr Asn Leu Pro Ser Glu Arg Ser Met Thr Gln Asn Gly
290 295 300

Tyr Ala Ala Ile Glu
305

<210> 7

<211> 1183

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (3)..(902)

<223>

<400> 7

ct ttc ttc ttc ttc gct tca gcc atg gga acc cta tct ctc tca tct 47
Phe Phe Phe Phe Ala Ser Ala Met Gly Thr Leu Ser Leu Ser Ser
1 5 10 15

tct ctc aaa cct tca ctc gtt tca tca aga ctc aat tca tct tcc tcc 95

```
Ser Leu Lys Pro Ser Leu Val Ser Ser Arg Leu Asn Ser Ser Ser Ser
                20                  25                  30

gca tct tct tct tcg ttt cct aaa cca aac aat ctc tac ctc aaa ccc   143
Ala Ser Ser Ser Ser Phe Pro Lys Pro Asn Asn Leu Tyr Leu Lys Pro
            35                  40                  45

acc aaa ctc att tca cca cct ctc aga aca act tca cca tcg cca ttg   191
Thr Lys Leu Ile Ser Pro Pro Leu Arg Thr Thr Ser Pro Ser Pro Leu
        50                  55                  60

aga ttc gcc aat gct tca atc gag atg tcg cag aca cag gaa tca gct   239
Arg Phe Ala Asn Ala Ser Ile Glu Met Ser Gln Thr Gln Glu Ser Ala
    65                  70                  75

att cgc gga gct gaa tct gac gtc atg ggt ctt ctc ctt agg gaa cga   287
Ile Arg Gly Ala Glu Ser Asp Val Met Gly Leu Leu Leu Arg Glu Arg
80                  85                  90                  95

atc gtc ttt ctc ggt agt agt atc gac gat ttc gtc gct gat gct att   335
Ile Val Phe Leu Gly Ser Ser Ile Asp Asp Phe Val Ala Asp Ala Ile
                100                 105                 110

atg agt cag ttg ctt ctc tta gat gct aaa gat cca aag aaa gat atc   383
Met Ser Gln Leu Leu Leu Leu Asp Ala Lys Asp Pro Lys Lys Asp Ile
            115                 120                 125

aaa ctc ttt atc aat tct cct ggt ggt tct ctc agt gca acc atg gct   431
Lys Leu Phe Ile Asn Ser Pro Gly Gly Ser Leu Ser Ala Thr Met Ala
        130                 135                 140

ata tac gat gtg gtt cag ctt gtg aga gct gat gtt tcg acg att gct   479
Ile Tyr Asp Val Val Gln Leu Val Arg Ala Asp Val Ser Thr Ile Ala
    145                 150                 155

ctt ggc att gct gca tca aca gct tcg att att ctt ggt gcg gga act   527
Leu Gly Ile Ala Ala Ser Thr Ala Ser Ile Ile Leu Gly Ala Gly Thr
160                 165                 170                 175

aaa ggc aag cgc ttt gct atg ccc aac acg agg ata atg att cat caa   575
Lys Gly Lys Arg Phe Ala Met Pro Asn Thr Arg Ile Met Ile His Gln
                180                 185                 190

cct ctt gga ggt gca agc ggt caa gct ata gat gtt gag att caa gct   623
Pro Leu Gly Gly Ala Ser Gly Gln Ala Ile Asp Val Glu Ile Gln Ala
            195                 200                 205

aag gaa gtt atg cat aac aag aac aat gtc acc agc att atc gcg gga   671
Lys Glu Val Met His Asn Lys Asn Asn Val Thr Ser Ile Ile Ala Gly
        210                 215                 220

tgt act agt cga tcg ttt gag cag gtt ctg aaa gat att gat agg gac   719
Cys Thr Ser Arg Ser Phe Glu Gln Val Leu Lys Asp Ile Asp Arg Asp
    225                 230                 235

cgg tac atg tct cca att gaa gca gtt gag tat ggt tta att gat gga   767
Arg Tyr Met Ser Pro Ile Glu Ala Val Glu Tyr Gly Leu Ile Asp Gly
240                 245                 250                 255

gtt att gat gga gac agc att att cct ctt gaa cct gtt cct gat aga   815
Val Ile Asp Gly Asp Ser Ile Ile Pro Leu Glu Pro Val Pro Asp Arg
                260                 265                 270

gtg aaa ccg aga gta aac tac gag gag att agc aag gat ccg atg aaa   863
```

Val Lys Pro Arg Val Asn Tyr Glu Glu Ile Ser Lys Asp Pro Met Lys
275 280 285

ttc ttg act ccc gag ata cct gat gat gag atc tac taa agccaagctc 912
Phe Leu Thr Pro Glu Ile Pro Asp Asp Glu Ile Tyr
290 295

gtctagaagc agggatcttc aaatgtgact aagactagca gtttcgagga aaagctcaat 972

ttcttctgcg gttactggta ttggctttgc gaaaccgaag ctggtagtac ttggcttttg 1032

tatctcatat ttcagttgtt cagaaaataa ttgttcttta aatcactctg ttttgaggaa 1092

aatgacttaa agaagctgta gttatctcgt ttatgacaat cccttcaagt gtttaatgga 1152

ttcaagaagt atcagtcagt atttttgtgg t 1183

<210> 8

<211> 299

<212> PRT

<213> Arabidopsis thaliana

<400> 8

Phe Phe Phe Phe Ala Ser Ala Met Gly Thr Leu Ser Leu Ser Ser Ser
1 5 10 15

Leu Lys Pro Ser Leu Val Ser Ser Arg Leu Asn Ser Ser Ser Ser Ala
20 25 30

Ser Ser Ser Ser Phe Pro Lys Pro Asn Asn Leu Tyr Leu Lys Pro Thr
35 40 45

Lys Leu Ile Ser Pro Pro Leu Arg Thr Thr Ser Pro Ser Pro Leu Arg
50 55 60

Phe Ala Asn Ala Ser Ile Glu Met Ser Gln Thr Gln Glu Ser Ala Ile
65 70 75 80

Arg Gly Ala Glu Ser Asp Val Met Gly Leu Leu Leu Arg Glu Arg Ile
85 90 95

Val Phe Leu Gly Ser Ser Ile Asp Asp Phe Val Ala Asp Ala Ile Met
100 105 110

Ser Gln Leu Leu Leu Leu Asp Ala Lys Asp Pro Lys Lys Asp Ile Lys
115 120 125

Leu Phe Ile Asn Ser Pro Gly Gly Ser Leu Ser Ala Thr Met Ala Ile
130 135 140

Tyr Asp Val Val Gln Leu Val Arg Ala Asp Val Ser Thr Ile Ala Leu
145 150 155 160

Gly Ile Ala Ala Ser Thr Ala Ser Ile Ile Leu Gly Ala Gly Thr Lys
165 170 175

Gly Lys Arg Phe Ala Met Pro Asn Thr Arg Ile Met Ile His Gln Pro
180 185 190

Leu Gly Gly Ala Ser Gly Gln Ala Ile Asp Val Glu Ile Gln Ala Lys
195 200 205

Glu Val Met His Asn Lys Asn Asn Val Thr Ser Ile Ile Ala Gly Cys
210 215 220

Thr Ser Arg Ser Phe Glu Gln Val Leu Lys Asp Ile Asp Arg Asp Arg
225 230 235 240

Tyr Met Ser Pro Ile Glu Ala Val Glu Tyr Gly Leu Ile Asp Gly Val
245 250 255

Ile Asp Gly Asp Ser Ile Ile Pro Leu Glu Pro Val Pro Asp Arg Val
260 265 270

Lys Pro Arg Val Asn Tyr Glu Glu Ile Ser Lys Asp Pro Met Lys Phe
275 280 285

Leu Thr Pro Glu Ile Pro Asp Asp Glu Ile Tyr
290 295

<210> 9

<211> 1056

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (61)..(876)

<223>

<400> 9
gagtaattta gcatctatcc acgcctgaac ccgaaaaact ctgaaagctg agctctggta 60

```
atg gcg ggt tta gca att tca cct cct ctc ggt ctt tcc ttc tct tct      108
Met Ala Gly Leu Ala Ile Ser Pro Pro Leu Gly Leu Ser Phe Ser Ser
1               5                  10                  15

cga act cga aac cct aaa ccc act tcc ttt cta tct cac aat caa agg      156
Arg Thr Arg Asn Pro Lys Pro Thr Ser Phe Leu Ser His Asn Gln Arg
            20                  25                  30

aat cct ata aga cgt ata gtt tct gct cta cag agt cca tat gga gat      204
Asn Pro Ile Arg Arg Ile Val Ser Ala Leu Gln Ser Pro Tyr Gly Asp
        35                  40                  45

tct ctg aaa gct gga ctt tct agt aat gtt tct gga tcc cca ata aag      252
Ser Leu Lys Ala Gly Leu Ser Ser Asn Val Ser Gly Ser Pro Ile Lys
    50                  55                  60

att gac aac aag gct cca aga ttt gga gtg ata gag gcg aaa aag gga      300
Ile Asp Asn Lys Ala Pro Arg Phe Gly Val Ile Glu Ala Lys Lys Gly
65                  70                  75                  80

aac ccc cca gta atg cct tca gtg atg acc cct gga gga cct tta gac      348
Asn Pro Pro Val Met Pro Ser Val Met Thr Pro Gly Gly Pro Leu Asp
                85                  90                  95

ctc tct tct gtg tta ttc cgt aac cgc ata atc ttc atc ggg caa cca      396
Leu Ser Ser Val Leu Phe Arg Asn Arg Ile Ile Phe Ile Gly Gln Pro
            100                 105                 110

att aac gca cag gtt gct cag cga gtc ata tct cag ctt gta acc ctt      444
Ile Asn Ala Gln Val Ala Gln Arg Val Ile Ser Gln Leu Val Thr Leu
        115                 120                 125

gca tct att gat gat aaa tcc gac atc ctg atg tac ttg aat tgt ccc      492
Ala Ser Ile Asp Asp Lys Ser Asp Ile Leu Met Tyr Leu Asn Cys Pro
    130                 135                 140

ggt ggc agt act tac tcc gtc cta aca att tat gac tgt atg tct tgg      540
Gly Gly Ser Thr Tyr Ser Val Leu Thr Ile Tyr Asp Cys Met Ser Trp
145                 150                 155                 160

ata aag cct aaa gtt gga aca gtg gcg ttt gga gta gct gca agc caa      588
Ile Lys Pro Lys Val Gly Thr Val Ala Phe Gly Val Ala Ala Ser Gln
                165                 170                 175

gga gca ttt ttt ctt gct gga ggt gaa aaa gga atg cgt tat gca atg      636
Gly Ala Phe Phe Leu Ala Gly Gly Glu Lys Gly Met Arg Tyr Ala Met
            180                 185                 190

cca aat act cgt gtc atg ata cat caa cca caa act gga tgc gga gga      684
Pro Asn Thr Arg Val Met Ile His Gln Pro Gln Thr Gly Cys Gly Gly
        195                 200                 205

cat gta gag gac gtg agg aga cag gtc aat gaa gcc atc gaa gcc cga      732
His Val Glu Asp Val Arg Arg Gln Val Asn Glu Ala Ile Glu Ala Arg
    210                 215                 220

caa aaa att gac agg atg tat gca gct ttc act gga caa cct ctg gag      780
Gln Lys Ile Asp Arg Met Tyr Ala Ala Phe Thr Gly Gln Pro Leu Glu
225                 230                 235                 240

aaa gtg cag caa tac act gaa aga gat cgt ttc tta tca gca tct gag      828
Lys Val Gln Gln Tyr Thr Glu Arg Asp Arg Phe Leu Ser Ala Ser Glu
                245                 250                 255
```

gcg ttt gag ttc ggg ctc att gat ggt cta ttg gaa aca gaa tac tga 876
Ala Phe Glu Phe Gly Leu Ile Asp Gly Leu Leu Glu Thr Glu Tyr
260 265 270

agcagcatac aggacaatgc acaacaacag ctcattgcaa tgttcaaagc ttccattttc 936

atttgaatat gaacggttgt aactgatatt tgtgcataaa tcagtttggt tttcttggtt 996

ttattgtcta ctaaacagaa tgagaaaact aaactgttta ttttttttact gaaaaatctg 1056

<210> 10

<211> 271

<212> PRT

<213> Arabidopsis thaliana

<400> 10

Met Ala Gly Leu Ala Ile Ser Pro Pro Leu Gly Leu Ser Phe Ser Ser
1 5 10 15

Arg Thr Arg Asn Pro Lys Pro Thr Ser Phe Leu Ser His Asn Gln Arg
20 25 30

Asn Pro Ile Arg Arg Ile Val Ser Ala Leu Gln Ser Pro Tyr Gly Asp
35 40 45

Ser Leu Lys Ala Gly Leu Ser Ser Asn Val Ser Gly Ser Pro Ile Lys
50 55 60

Ile Asp Asn Lys Ala Pro Arg Phe Gly Val Ile Glu Ala Lys Lys Gly
65 70 75 80

Asn Pro Pro Val Met Pro Ser Val Met Thr Pro Gly Gly Pro Leu Asp
85 90 95

Leu Ser Ser Val Leu Phe Arg Asn Arg Ile Ile Phe Ile Gly Gln Pro
100 105 110

Ile Asn Ala Gln Val Ala Gln Arg Val Ile Ser Gln Leu Val Thr Leu
115 120 125

Ala Ser Ile Asp Asp Lys Ser Asp Ile Leu Met Tyr Leu Asn Cys Pro
130 135 140

Gly Gly Ser Thr Tyr Ser Val Leu Thr Ile Tyr Asp Cys Met Ser Trp
145 150 155 160

Ile Lys Pro Lys Val Gly Thr Val Ala Phe Gly Val Ala Ala Ser Gln
165 170 175

Gly Ala Phe Phe Leu Ala Gly Gly Glu Lys Gly Met Arg Tyr Ala Met
180 185 190

Pro Asn Thr Arg Val Met Ile His Gln Pro Gln Thr Gly Cys Gly Gly
195 200 205

His Val Glu Asp Val Arg Arg Gln Val Asn Glu Ala Ile Glu Ala Arg
210 215 220

Gln Lys Ile Asp Arg Met Tyr Ala Ala Phe Thr Gly Gln Pro Leu Glu
225 230 235 240

Lys Val Gln Gln Tyr Thr Glu Arg Asp Arg Phe Leu Ser Ala Ser Glu
245 250 255

Ala Phe Glu Phe Gly Leu Ile Asp Gly Leu Leu Glu Thr Glu Tyr
260 265 270

<210> 11

<211> 1448

<212> DNA

<213> Nicotiana tabacum

<220>

<221> CDS

<222> (2)..(1162)

<223>

<400> 11

g cgg ccg ctg gct tct tct ttg ctt ctc tct ccg ctt tct agc tcg acg 49
Arg Pro Leu Ala Ser Ser Leu Leu Leu Ser Pro Leu Ser Ser Ser Thr
1 5 10 15

gtt act gaa aat cgc gag ctg ggt tct ggt aaa tca act ttc ata tcc 97
Val Thr Glu Asn Arg Glu Leu Gly Ser Gly Lys Ser Thr Phe Ile Ser
20 25 30

agt ccc aat ttc tcc ttt gca act tct gtt cac agt tgc agg cca aac 145
Ser Pro Asn Phe Ser Phe Ala Thr Ser Val His Ser Cys Arg Pro Asn
35 40 45

ggc gtt cga ggt tat tgt tac agg tct ccg gta gct aag tct ttg gac 193
Gly Val Arg Gly Tyr Cys Tyr Arg Ser Pro Val Ala Lys Ser Leu Asp
50 55 60

```
cat ata ccc caa aaa ttc aga ctg gaa aat ctc aaa gat gga cta ctg      241
His Ile Pro Gln Lys Phe Arg Leu Glu Asn Leu Lys Asp Gly Leu Leu
65                  70                  75                  80

gac aac tat aaa agt gcc cct cag tat ctt tac ggc ctt agt cct tca      289
Asp Asn Tyr Lys Ser Ala Pro Gln Tyr Leu Tyr Gly Leu Ser Pro Ser
                85                  90                  95

cag atg gat atg ttc atg aca gaa gat aac cca gta cgg cga cag tca      337
Gln Met Asp Met Phe Met Thr Glu Asp Asn Pro Val Arg Arg Gln Ser
            100                 105                 110

gaa agt gcc act gag gat agt ata tct tca gcc aat aac tat ctg agc      385
Glu Ser Ala Thr Glu Asp Ser Ile Ser Ser Ala Asn Asn Tyr Leu Ser
        115                 120                 125

aat ggt gga atg tgg agt atg tcc ggc atg aac gat cgg ggc ccc tcg      433
Asn Gly Gly Met Trp Ser Met Ser Gly Met Asn Asp Arg Gly Pro Ser
    130                 135                 140

aaa tac agt atg agt gtc agc atg tac cgt gga gga aca aga gga tct      481
Lys Tyr Ser Met Ser Val Ser Met Tyr Arg Gly Gly Thr Arg Gly Ser
145                 150                 155                 160

gga aga cct cga act gcg cct cct gat ttg cca tct ttg ctt ttg gat      529
Gly Arg Pro Arg Thr Ala Pro Pro Asp Leu Pro Ser Leu Leu Leu Asp
                165                 170                 175

gct cga att gtc tat ctg ggc atg cct att gta cca gct gtt aca gag      577
Ala Arg Ile Val Tyr Leu Gly Met Pro Ile Val Pro Ala Val Thr Glu
            180                 185                 190

ctt ctt gtt gct cag ttt atg tgg ttg gat tat gac aat cca tca aag      625
Leu Leu Val Ala Gln Phe Met Trp Leu Asp Tyr Asp Asn Pro Ser Lys
        195                 200                 205

cct ata tac cta tat ata aac tca tca ggc aca cag aat gag aag atg      673
Pro Ile Tyr Leu Tyr Ile Asn Ser Ser Gly Thr Gln Asn Glu Lys Met
    210                 215                 220

gag act gtt ggg tct gaa aca gag gca tat gcc atc gct gac aca atg      721
Glu Thr Val Gly Ser Glu Thr Glu Ala Tyr Ala Ile Ala Asp Thr Met
225                 230                 235                 240

gca tac tgc aaa tca gat atc tat aca gtg aac tgt ggc atg gca tat      769
Ala Tyr Cys Lys Ser Asp Ile Tyr Thr Val Asn Cys Gly Met Ala Tyr
                245                 250                 255

ggt caa gca gca atg ctt ctg tca ctg gga aag aag ggg ttc cgt gct      817
Gly Gln Ala Ala Met Leu Leu Ser Leu Gly Lys Lys Gly Phe Arg Ala
            260                 265                 270

atg cag cca aat tca tct aca aaa ttg tat tta cct aaa gtc agc aaa      865
Met Gln Pro Asn Ser Ser Thr Lys Leu Tyr Leu Pro Lys Val Ser Lys
        275                 280                 285

tcc agt gga gca gtg ata gat atg tgg atc agg gcc aaa gaa cta gat      913
Ser Ser Gly Ala Val Ile Asp Met Trp Ile Arg Ala Lys Glu Leu Asp
    290                 295                 300

gca aac act gag tat tac ctt gaa cta tta gcg aaa gga gtt gga aaa      961
Ala Asn Thr Glu Tyr Tyr Leu Glu Leu Leu Ala Lys Gly Val Gly Lys
305                 310                 315                 320
```

cca aag gaa gaa atc gag aaa gat att caa cgc cct aaa tat ctg cgg 1009
Pro Lys Glu Glu Ile Glu Lys Asp Ile Gln Arg Pro Lys Tyr Leu Arg
325 330 335

gca caa gaa gcc att gac tat ggc att gcg gac aag ata atc gat tca 1057
Ala Gln Glu Ala Ile Asp Tyr Gly Ile Ala Asp Lys Ile Ile Asp Ser
340 345 350

aga gac aat gca ttt gag aaa agg aac tat ggt gag ata ctc gcc caa 1105
Arg Asp Asn Ala Phe Glu Lys Arg Asn Tyr Gly Glu Ile Leu Ala Gln
355 360 365

tct aga gct atg agg aaa gcc gga cca ggt gct cag gct gct cca tct 1153
Ser Arg Ala Met Arg Lys Ala Gly Pro Gly Ala Gln Ala Ala Pro Ser
370 375 380

ggc tcc agg tgactggaag agcggtaatg gtcccaagct ttcaggaaca 1202
Gly Ser Arg
385

actgttgttc ccttatagtt tcgaggaaca aagttgctgg ttacttggtc tgtgccggta 1262

taatgtaact gggacaaaga acatattgta gaaaccttgt ttgagctgtg aagtataggg 1322

gttttacaac tattatgcac aggtctgcaa agagtaccca taatgtcaat tggttgtacc 1382

agtatcaaac aatcagatag tgccagtgta tggtataaat gaatatagat ctctctgagc 1442

ggccgc 1448

<210> 12

<211> 387

<212> PRT

<213> Nicotiana tabacum

<400> 12

Arg Pro Leu Ala Ser Ser Leu Leu Leu Ser Pro Leu Ser Ser Ser Thr
1 5 10 15

Val Thr Glu Asn Arg Glu Leu Gly Ser Gly Lys Ser Thr Phe Ile Ser
20 25 30

Ser Pro Asn Phe Ser Phe Ala Thr Ser Val His Ser Cys Arg Pro Asn
35 40 45

Gly Val Arg Gly Tyr Cys Tyr Arg Ser Pro Val Ala Lys Ser Leu Asp
50 55 60

His Ile Pro Gln Lys Phe Arg Leu Glu Asn Leu Lys Asp Gly Leu Leu
65 70 75 80

```
Asp Asn Tyr Lys Ser Ala Pro Gln Tyr Leu Tyr Gly Leu Ser Pro Ser
                85                  90                  95

Gln Met Asp Met Phe Met Thr Glu Asp Asn Pro Val Arg Arg Gln Ser
            100                 105                 110

Glu Ser Ala Thr Glu Asp Ser Ile Ser Ser Ala Asn Asn Tyr Leu Ser
        115                 120                 125

Asn Gly Gly Met Trp Ser Met Ser Gly Met Asn Asp Arg Gly Pro Ser
    130                 135                 140

Lys Tyr Ser Met Ser Val Ser Met Tyr Arg Gly Gly Thr Arg Gly Ser
145                 150                 155                 160

Gly Arg Pro Arg Thr Ala Pro Pro Asp Leu Pro Ser Leu Leu Leu Asp
                165                 170                 175

Ala Arg Ile Val Tyr Leu Gly Met Pro Ile Val Pro Ala Val Thr Glu
            180                 185                 190

Leu Leu Val Ala Gln Phe Met Trp Leu Asp Tyr Asp Asn Pro Ser Lys
        195                 200                 205

Pro Ile Tyr Leu Tyr Ile Asn Ser Ser Gly Thr Gln Asn Glu Lys Met
    210                 215                 220

Glu Thr Val Gly Ser Glu Thr Glu Ala Tyr Ala Ile Ala Asp Thr Met
225                 230                 235                 240

Ala Tyr Cys Lys Ser Asp Ile Tyr Thr Val Asn Cys Gly Met Ala Tyr
                245                 250                 255

Gly Gln Ala Ala Met Leu Leu Ser Leu Gly Lys Lys Gly Phe Arg Ala
            260                 265                 270

Met Gln Pro Asn Ser Ser Thr Lys Leu Tyr Leu Pro Lys Val Ser Lys
        275                 280                 285

Ser Ser Gly Ala Val Ile Asp Met Trp Ile Arg Ala Lys Glu Leu Asp
    290                 295                 300

Ala Asn Thr Glu Tyr Tyr Leu Glu Leu Leu Ala Lys Gly Val Gly Lys
305                 310                 315                 320

Pro Lys Glu Glu Ile Glu Lys Asp Ile Gln Arg Pro Lys Tyr Leu Arg
                325                 330                 335
```

Ala Gln Glu Ala Ile Asp Tyr Gly Ile Ala Asp Lys Ile Ile Asp Ser
340 345 350

Arg Asp Asn Ala Phe Glu Lys Arg Asn Tyr Gly Glu Ile Leu Ala Gln
355 360 365

Ser Arg Ala Met Arg Lys Ala Gly Pro Gly Ala Gln Ala Ala Pro Ser
370 375 380

Gly Ser Arg
385

<210> 13

<211> 1246

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (38)..(1030)

<223>

<400> 13
attttcgcga gcttccgtgt ccaagagctc ctcgacc atg gcg tct tgt tta caa 55
Met Ala Ser Cys Leu Gln
1 5

gca tcc atg aat tct ctg ctt cca cgc tct tct tct ttt tct cct cat 103
Ala Ser Met Asn Ser Leu Leu Pro Arg Ser Ser Ser Phe Ser Pro His
10 15 20

cct cct cta tct tcg aat tca tcc ggg aga aga aac ttg aag act ttt 151
Pro Pro Leu Ser Ser Asn Ser Ser Gly Arg Arg Asn Leu Lys Thr Phe
25 30 35

cgt tac gcc ttt cgc gcc aaa gcc tct gcc aaa atc cct atg cct ccg 199
Arg Tyr Ala Phe Arg Ala Lys Ala Ser Ala Lys Ile Pro Met Pro Pro
40 45 50

ata aat cca aag gat cct ttc ctc tcc acg ctc gct tct att gcc gcg 247
Ile Asn Pro Lys Asp Pro Phe Leu Ser Thr Leu Ala Ser Ile Ala Ala
55 60 65 70

aat tct ccg gaa aag ctt ctc aat cgg ccg gtt aac gct gat gtg ccg 295
Asn Ser Pro Glu Lys Leu Leu Asn Arg Pro Val Asn Ala Asp Val Pro
75 80 85

cca tat ctt gac atc ttt gac tcc cct cag ctc atg tct tct cct gca 343
Pro Tyr Leu Asp Ile Phe Asp Ser Pro Gln Leu Met Ser Ser Pro Ala
90 95 100

```
cag gtt gaa aga tca gtg gct tat aac gag cac cga ccg aga act cct      391
Gln Val Glu Arg Ser Val Ala Tyr Asn Glu His Arg Pro Arg Thr Pro
        105                 110                 115

cca cca gac ttg cca tct atg ctt ctt gac ggg aga att gtt tac att      439
Pro Pro Asp Leu Pro Ser Met Leu Leu Asp Gly Arg Ile Val Tyr Ile
    120                 125                 130

gga atg cct ctt gtg ccg gca gtg act gag cta gtt gtc gct gag cta      487
Gly Met Pro Leu Val Pro Ala Val Thr Glu Leu Val Val Ala Glu Leu
135                 140                 145                 150

atg tat ctt cag tgg ctg gat ccc aag gaa ccc att tac att tac atc      535
Met Tyr Leu Gln Trp Leu Asp Pro Lys Glu Pro Ile Tyr Ile Tyr Ile
                155                 160                 165

aac tcc aca ggg acc act cgt gat gat gga gag acg gtt gga atg gaa      583
Asn Ser Thr Gly Thr Thr Arg Asp Asp Gly Glu Thr Val Gly Met Glu
            170                 175                 180

tca gaa ggg ttt gcg atc tat gac tct ttg atg caa ctt aaa aac gag      631
Ser Glu Gly Phe Ala Ile Tyr Asp Ser Leu Met Gln Leu Lys Asn Glu
        185                 190                 195

gta cat aca gta tgt gtg gga gca gcc ata ggt cag gcc tgt cta tta      679
Val His Thr Val Cys Val Gly Ala Ala Ile Gly Gln Ala Cys Leu Leu
    200                 205                 210

ctt tct gcg gga acc aag ggt aaa cgg ttt atg atg cca cat gcc aaa      727
Leu Ser Ala Gly Thr Lys Gly Lys Arg Phe Met Met Pro His Ala Lys
215                 220                 225                 230

gcg atg att cag caa cca cgt gta cct tct tct ggg ttg atg cct gcc      775
Ala Met Ile Gln Gln Pro Arg Val Pro Ser Ser Gly Leu Met Pro Ala
                235                 240                 245

agt gat gtc ctg att cgg gcc aaa gag gtc att aca aat agg gat ata      823
Ser Asp Val Leu Ile Arg Ala Lys Glu Val Ile Thr Asn Arg Asp Ile
            250                 255                 260

ctt gtg gaa cta cta tca aag cat act ggg aat tcc gtg gag act gta      871
Leu Val Glu Leu Leu Ser Lys His Thr Gly Asn Ser Val Glu Thr Val
        265                 270                 275

gct aac gta atg aga agg cca tat tac atg gat gca cca aaa gct aaa      919
Ala Asn Val Met Arg Arg Pro Tyr Tyr Met Asp Ala Pro Lys Ala Lys
    280                 285                 290

gaa ttt gga gtc att gac agg att ctt tgg cgc ggt caa gaa aag att      967
Glu Phe Gly Val Ile Asp Arg Ile Leu Trp Arg Gly Gln Glu Lys Ile
295                 300                 305                 310

att gcg gac gtg gtt cct tca gag gaa ttc gac aag aat gca ggg att     1015
Ile Ala Asp Val Val Pro Ser Glu Glu Phe Asp Lys Asn Ala Gly Ile
                315                 320                 325

aaa agc gta gta tga gtctagtctt aagttttctt ggcctaaatc atactgcgtc     1070
Lys Ser Val Val
            330

atggagaaga acaaatagac tgaccaaaat cacattggcc gcagactgcc ttgtttcaaa   1130

tcacttggta aatgtgaaca tgcgattagg agaatcatac ttaaaggatc ttgaaatatt   1190
```

atgataaaat tgtaatgtgt ttgttcgtta gcaatagtaa atacaatctt caactc 1246

<210> 14

<211> 330

<212> PRT

<213> Arabidopsis thaliana

<400> 14

Met Ala Ser Cys Leu Gln Ala Ser Met Asn Ser Leu Leu Pro Arg Ser
1 5 10 15

Ser Ser Phe Ser Pro His Pro Pro Leu Ser Ser Asn Ser Ser Gly Arg
20 25 30

Arg Asn Leu Lys Thr Phe Arg Tyr Ala Phe Arg Ala Lys Ala Ser Ala
35 40 45

Lys Ile Pro Met Pro Pro Ile Asn Pro Lys Asp Pro Phe Leu Ser Thr
50 55 60

Leu Ala Ser Ile Ala Ala Asn Ser Pro Glu Lys Leu Leu Asn Arg Pro
65 70 75 80

Val Asn Ala Asp Val Pro Pro Tyr Leu Asp Ile Phe Asp Ser Pro Gln
85 90 95

Leu Met Ser Ser Pro Ala Gln Val Glu Arg Ser Val Ala Tyr Asn Glu
100 105 110

His Arg Pro Arg Thr Pro Pro Pro Asp Leu Pro Ser Met Leu Leu Asp
115 120 125

Gly Arg Ile Val Tyr Ile Gly Met Pro Leu Val Pro Ala Val Thr Glu
130 135 140

Leu Val Val Ala Glu Leu Met Tyr Leu Gln Trp Leu Asp Pro Lys Glu
145 150 155 160

Pro Ile Tyr Ile Tyr Ile Asn Ser Thr Gly Thr Thr Arg Asp Asp Gly
165 170 175

Glu Thr Val Gly Met Glu Ser Glu Gly Phe Ala Ile Tyr Asp Ser Leu
180 185 190

Met Gln Leu Lys Asn Glu Val His Thr Val Cys Val Gly Ala Ala Ile
195 200 205

Gly Gln Ala Cys Leu Leu Leu Ser Ala Gly Thr Lys Gly Lys Arg Phe
210 215 220

Met Met Pro His Ala Lys Ala Met Ile Gln Gln Pro Arg Val Pro Ser
225 230 235 240

Ser Gly Leu Met Pro Ala Ser Asp Val Leu Ile Arg Ala Lys Glu Val
245 250 255

Ile Thr Asn Arg Asp Ile Leu Val Glu Leu Leu Ser Lys His Thr Gly
260 265 270

Asn Ser Val Glu Thr Val Ala Asn Val Met Arg Arg Pro Tyr Tyr Met
275 280 285

Asp Ala Pro Lys Ala Lys Glu Phe Gly Val Ile Asp Arg Ile Leu Trp
290 295 300

Arg Gly Gln Glu Lys Ile Ile Ala Asp Val Val Pro Ser Glu Glu Phe
305 310 315 320

Asp Lys Asn Ala Gly Ile Lys Ser Val Val
325 330

<210> 15

<211> 1236

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (66)..(983)

<223>

<400> 15
agatcgttat cgtttcgggg tcacagggac tttcactctt tctctctctc tgcaacaaag 60

aagaa atg gag gta gca gca gcg act gcg acg agc ttc aca acg ctt cga 110
Met Glu Val Ala Ala Ala Thr Ala Thr Ser Phe Thr Thr Leu Arg
1 5 10 15

gct cgt acg tca gcg att atc ccg tct tct aca cgt aat ctg aga tct 158

```
Ala Arg Thr Ser Ala Ile Ile Pro Ser Ser Thr Arg Asn Leu Arg Ser
                  20                  25                  30

aaa ccg aga ttt tct tca tct tca tct ctc aga gct tct ctt tcg aat      206
Lys Pro Arg Phe Ser Ser Ser Ser Ser Leu Arg Ala Ser Leu Ser Asn
              35                  40                  45

ggc ttt ctt tcg ccg tat acc gga gga agc atc tct agt gac tta tgc      254
Gly Phe Leu Ser Pro Tyr Thr Gly Gly Ser Ile Ser Ser Asp Leu Cys
          50                  55                  60

ggc gct aag ctt cgt gcg gaa tcg ctt aat ccg tta aat ttt tcc agt      302
Gly Ala Lys Leu Arg Ala Glu Ser Leu Asn Pro Leu Asn Phe Ser Ser
      65                  70                  75

tcc aag cct aaa cgc gga gtt gtc act atg gtt ata cct ttc tca aag      350
Ser Lys Pro Lys Arg Gly Val Val Thr Met Val Ile Pro Phe Ser Lys
80                  85                  90                  95

gga agt gca cac gaa caa cct cct cct gat ttg gca tca tat ttg ttc      398
Gly Ser Ala His Glu Gln Pro Pro Pro Asp Leu Ala Ser Tyr Leu Phe
                  100                 105                 110

aag aac cga att gta tat ttg gga atg tct ctc gta cct tca gtt act      446
Lys Asn Arg Ile Val Tyr Leu Gly Met Ser Leu Val Pro Ser Val Thr
              115                 120                 125

gag ttg ata ctt gcg gag ttt ctt tac ctt cag tat gaa gac gag gaa      494
Glu Leu Ile Leu Ala Glu Phe Leu Tyr Leu Gln Tyr Glu Asp Glu Glu
          130                 135                 140

aag cct att tac ctt tac ata aac tcg act ggg aca acc aag aat ggt      542
Lys Pro Ile Tyr Leu Tyr Ile Asn Ser Thr Gly Thr Thr Lys Asn Gly
      145                 150                 155

gaa aag ttg ggc tat gat act gag gct ttt gca atc tat gat gtc atg      590
Glu Lys Leu Gly Tyr Asp Thr Glu Ala Phe Ala Ile Tyr Asp Val Met
160                 165                 170                 175

ggg tat gtc aaa cca cca atc ttt act ctt tgc gtc ggg aat gcg tgg      638
Gly Tyr Val Lys Pro Pro Ile Phe Thr Leu Cys Val Gly Asn Ala Trp
                  180                 185                 190

ggt gaa gct gct ttg ctt ctg act gct ggt gca aaa gga aat cga tct      686
Gly Glu Ala Ala Leu Leu Leu Thr Ala Gly Ala Lys Gly Asn Arg Ser
              195                 200                 205

gcg ttg ccc tca tca act att atg ata aag cag ccc att gct cga ttt      734
Ala Leu Pro Ser Ser Thr Ile Met Ile Lys Gln Pro Ile Ala Arg Phe
          210                 215                 220

caa ggc caa gca act gat gtt gaa att gca agg aaa gaa atc aag cac      782
Gln Gly Gln Ala Thr Asp Val Glu Ile Ala Arg Lys Glu Ile Lys His
      225                 230                 235

ata aag aca gaa atg gtc aag ctg tat tca aag cat att ggt aaa tcc      830
Ile Lys Thr Glu Met Val Lys Leu Tyr Ser Lys His Ile Gly Lys Ser
240                 245                 250                 255

ccg gag cag att gaa gct gac atg aaa cgc ccg aaa tat ttt agt ccc      878
Pro Glu Gln Ile Glu Ala Asp Met Lys Arg Pro Lys Tyr Phe Ser Pro
                  260                 265                 270

act gag gct gtt gaa tat ggg atc att gat aag gtg gtt tac aat gaa      926
```

Thr Glu Ala Val Glu Tyr Gly Ile Ile Asp Lys Val Val Tyr Asn Glu
275 280 285

agg ggc agc caa gac aga gga gtt gtg tct gac ctt aaa aag gca caa 974
Arg Gly Ser Gln Asp Arg Gly Val Val Ser Asp Leu Lys Lys Ala Gln
290 295 300

ctc att tga atgtcagaac tgtcttccga aatcccatga ttaacaggtt 1023
Leu Ile
305

ggagatctta ccgctgatca aatggggaat cagtgaacca ttcaccggca cagaactgag 1083

gtaaagtctg gaaaacatgt taaaaaaggt tactagtaat gctgcaattg tagggttatt 1143

tgaacagaaa caaacccata tgtgtaggct tgtgaatgcc tagaaacagg attggtgtat 1203

cttcaatata tgtttctaag atgaatcaat ttc 1236

<210> 16

<211> 305

<212> PRT

<213> Arabidopsis thaliana

<400> 16

Met Glu Val Ala Ala Ala Thr Ala Thr Ser Phe Thr Thr Leu Arg Ala
1 5 10 15

Arg Thr Ser Ala Ile Ile Pro Ser Ser Thr Arg Asn Leu Arg Ser Lys
20 25 30

Pro Arg Phe Ser Ser Ser Ser Ser Leu Arg Ala Ser Leu Ser Asn Gly
35 40 45

Phe Leu Ser Pro Tyr Thr Gly Gly Ser Ile Ser Ser Asp Leu Cys Gly
50 55 60

Ala Lys Leu Arg Ala Glu Ser Leu Asn Pro Leu Asn Phe Ser Ser Ser
65 70 75 80

Lys Pro Lys Arg Gly Val Val Thr Met Val Ile Pro Phe Ser Lys Gly
85 90 95

Ser Ala His Glu Gln Pro Pro Pro Asp Leu Ala Ser Tyr Leu Phe Lys
100 105 110

Asn Arg Ile Val Tyr Leu Gly Met Ser Leu Val Pro Ser Val Thr Glu
115 120 125

Leu Ile Leu Ala Glu Phe Leu Tyr Leu Gln Tyr Glu Asp Glu Glu Lys
130 135 140

Pro Ile Tyr Leu Tyr Ile Asn Ser Thr Gly Thr Thr Lys Asn Gly Glu
145 150 155 160

Lys Leu Gly Tyr Asp Thr Glu Ala Phe Ala Ile Tyr Asp Val Met Gly
165 170 175

Tyr Val Lys Pro Pro Ile Phe Thr Leu Cys Val Gly Asn Ala Trp Gly
180 185 190

Glu Ala Ala Leu Leu Leu Thr Ala Gly Ala Lys Gly Asn Arg Ser Ala
195 200 205

Leu Pro Ser Ser Thr Ile Met Ile Lys Gln Pro Ile Ala Arg Phe Gln
210 215 220

Gly Gln Ala Thr Asp Val Glu Ile Ala Arg Lys Glu Ile Lys His Ile
225 230 235 240

Lys Thr Glu Met Val Lys Leu Tyr Ser Lys His Ile Gly Lys Ser Pro
245 250 255

Glu Gln Ile Glu Ala Asp Met Lys Arg Pro Lys Tyr Phe Ser Pro Thr
260 265 270

Glu Ala Val Glu Tyr Gly Ile Ile Asp Lys Val Val Tyr Asn Glu Arg
275 280 285

Gly Ser Gln Asp Arg Gly Val Val Ser Asp Leu Lys Lys Ala Gln Leu
290 295 300

Ile
305

<210> 17

<211> 906

<212> DNA

<213> Nicotiana tabacum

<220>

<221> CDS

<222> (45)..(755)

<223>

<400> 17

```
gcggccgctc caagattcat ccccaactct caacacattc aact atg cgc acc caa      56
                                                 Met Arg Thr Gln
                                                 1

att gtt cac aaa ctc ttt aac cga aga atc aac gga acc cct ttg aat      104
Ile Val His Lys Leu Phe Asn Arg Arg Ile Asn Gly Thr Pro Leu Asn
5                   10                  15                  20

agt agt aag aga ttt tat ggg gta ata cca atg gta ata gag cac tct      152
Ser Ser Lys Arg Phe Tyr Gly Val Ile Pro Met Val Ile Glu His Ser
                25                  30                  35

tca aga gga gaa agg gct tat gac ata ttc tca agg cta tta aag gaa      200
Ser Arg Gly Glu Arg Ala Tyr Asp Ile Phe Ser Arg Leu Leu Lys Glu
            40                  45                  50

cga att att tgc att aac ggc ccc att gat gat tcc act tct cat gtt      248
Arg Ile Ile Cys Ile Asn Gly Pro Ile Asp Asp Ser Thr Ser His Val
        55                  60                  65

gtt gtt gct cag ctt ctt ttt ctt gaa tct gag aac cct tct aag cct      296
Val Val Ala Gln Leu Leu Phe Leu Glu Ser Glu Asn Pro Ser Lys Pro
    70                  75                  80

att cac aag tac ctc aac tct cca ggt ggc gct gtt aca gct ggt ctt      344
Ile His Lys Tyr Leu Asn Ser Pro Gly Gly Ala Val Thr Ala Gly Leu
85                  90                  95                  100

gca atc tat gat acc acg cag tat atc cga tct cca att cat act ata      392
Ala Ile Tyr Asp Thr Thr Gln Tyr Ile Arg Ser Pro Ile His Thr Ile
                105                 110                 115

tgc cta ggt caa gca gct tca atg gga tcc ctt ctc tta gct gca ggt      440
Cys Leu Gly Gln Ala Ala Ser Met Gly Ser Leu Leu Leu Ala Ala Gly
            120                 125                 130

gca aag ggt gag aga cga tct ctc cct aat gct tca gtt atg att cac      488
Ala Lys Gly Glu Arg Arg Ser Leu Pro Asn Ala Ser Val Met Ile His
        135                 140                 145

cag cct ttc ggt ggg tat agc ggg cag gct aaa gat ttg acg atc cac      536
Gln Pro Phe Gly Gly Tyr Ser Gly Gln Ala Lys Asp Leu Thr Ile His
    150                 155                 160

aca aaa cag ata gtt cgg gta tgg gat act ttg aat gac cta tat gca      584
Thr Lys Gln Ile Val Arg Val Trp Asp Thr Leu Asn Asp Leu Tyr Ala
165                 170                 175                 180

aag cat aca gga caa cct ata gaa ata att caa aag aat atg gat agg      632
Lys His Thr Gly Gln Pro Ile Glu Ile Ile Gln Lys Asn Met Asp Arg
                185                 190                 195

gat tat ttc atg aca cct gaa gag gcg aag gag ttt gga ata atc gat      680
Asp Tyr Phe Met Thr Pro Glu Glu Ala Lys Glu Phe Gly Ile Ile Asp
            200                 205                 210

gaa gtt ata gat gaa cga cca atg gct tta gta act gat gct gtt gca      728
Glu Val Ile Asp Glu Arg Pro Met Ala Leu Val Thr Asp Ala Val Ala
        215                 220                 225
```

aat gaa gcc aaa gaa aaa ggt tca agc tagaaaaatt gctgtaatac 775
Asn Glu Ala Lys Glu Lys Gly Ser Ser
230 235

tgatctcatt gcagtctttg ttagcattta ccatcgctaa ctagttctcc attttactta 835

ctggtgtatt tactttctag tattttattt gatgaggcga tacctcatta ctttgttttc 895

tcagcggccg c 906

<210> 18

<211> 237

<212> PRT

<213> Nicotiana tabacum

<400> 18

Met Arg Thr Gln Ile Val His Lys Leu Phe Asn Arg Arg Ile Asn Gly
1 5 10 15

Thr Pro Leu Asn Ser Ser Lys Arg Phe Tyr Gly Val Ile Pro Met Val
20 25 30

Ile Glu His Ser Ser Arg Gly Glu Arg Ala Tyr Asp Ile Phe Ser Arg
35 40 45

Leu Leu Lys Glu Arg Ile Ile Cys Ile Asn Gly Pro Ile Asp Asp Ser
50 55 60

Thr Ser His Val Val Val Ala Gln Leu Leu Phe Leu Glu Ser Glu Asn
65 70 75 80

Pro Ser Lys Pro Ile His Lys Tyr Leu Asn Ser Pro Gly Gly Ala Val
85 90 95

Thr Ala Gly Leu Ala Ile Tyr Asp Thr Thr Gln Tyr Ile Arg Ser Pro
100 105 110

Ile His Thr Ile Cys Leu Gly Gln Ala Ala Ser Met Gly Ser Leu Leu
115 120 125

Leu Ala Ala Gly Ala Lys Gly Glu Arg Arg Ser Leu Pro Asn Ala Ser
130 135 140

Val Met Ile His Gln Pro Phe Gly Gly Tyr Ser Gly Gln Ala Lys Asp
145 150 155 160

Leu Thr Ile His Thr Lys Gln Ile Val Arg Val Trp Asp Thr Leu Asn
165 170 175

Asp Leu Tyr Ala Lys His Thr Gly Gln Pro Ile Glu Ile Ile Gln Lys
180 185 190

Asn Met Asp Arg Asp Tyr Phe Met Thr Pro Glu Glu Ala Lys Glu Phe
195 200 205

Gly Ile Ile Asp Glu Val Ile Asp Glu Arg Pro Met Ala Leu Val Thr
210 215 220

Asp Ala Val Ala Asn Glu Ala Lys Glu Lys Gly Ser Ser
225 230 235

<210> 19

<211> 447

<212> DNA

<213> Nicotiana tabacum

<400> 19
gcggccgctt gcggacaaga taatcgattc aagagacaat gtatttgaga aaaggaacta 60

tgatgagata ctcgcccaat ctagagctat gaggaaagcc ggaccaggtg ctcaggctgc 120

tccatctggc ttcaggtgac tggaagagcg gtaatggtcc caaactttca ggaacaactg 180

ttgttccctt atagtttcga ggaacaaagt tgctggttac ttggtctgtg ccggtataat 240

gtaactggga caaagaacat attgtagaaa ccttgtttga gctgtgaagt ataggggttt 300

tacaactatt atgcacaggt ctgcaaagag tacccataat gtcaattggt tgtaccagta 360

tcaaacaatc agatagtgcc agtgtatggt ataaatgaat atagatctct ctgatgtcat 420

ttttctttta tcatgttcag cggccgc 447

<210> 20

<211> 996

<212> DNA

<213> Nicotiana tabacum

<400> 20
gcggccgctt gcggacaaga taatcgattc aagagacaat gtatttgaga aaaggaacta 60

tgatgagata ctcgcccaat ctagagctat gaggaaagcc ggaccaggtg ctcaggctgc 120

tccatctggc ttcaggtgac tggaagagcg gtaatggtcc caaactttca ggaacaactg 180

ttgttccctt atagtttcga ggaacaaagt tgctggttac ttggtctgtg ccggtataat 240

gtaactggga caaagaacat attgtagaaa ccttgtttga gctgtgaagt ataggggttt 300

tacaactatt atgcacaggt ctgcaaagag tacccataat gtcaattggt tgtaccaggc 360

ggccgctggc ttcttctttg cttctctctc cgctttctag ctcgacggtt actgaaaatc 420

gcgagctggg ttctggtaaa tcaactttca tatccagtcc caatttctcc tttgcaactt 480

ctgttcacag ttgcaggcca aacggcgttc gaggttattg ttacaggtct ccggtagcta 540

agtctttgga ccatataccc caaaaattca gactggaaaa tctcaaagat ggactactgg 600

acaactataa aagtgcccct cagtatcttt acggccttag tccttcacag atggatatgt 660

tcatgacaga agataaccca gtacggcgac agtcagaaag tgccactgag gatagtatat 720

ctgcggccgc tggcagatgc tccacgaarg gataccacag caggtcttgg ttagtccata 780

cacatcgtat aatttatggc tgatagtggt tgtacgactt gcagtgttat tttgcaattt 840

cttttgttta atctacatat tgaactcttt tgatctactt attcaaaaac atgaaatcct 900

gagcagacta gatgcatttg tttaatatca tgaatgcaag gaatccacct acagctgata 960

tgtatacaaa gataccttt tttcaagagc ggccgc 996

<210> 21

<211> 602

<212> DNA

<213> Nicotiana tabacum

<220>

<221> CDS

<222> (2)..(193)

<223>

<400> 21

g cgg ccg ctg gaa gat gtg cgg cgc caa gtg aac gaa gcg gtt caa cct 49
Arg Pro Leu Glu Asp Val Arg Arg Gln Val Asn Glu Ala Val Gln Pro
1 5 10 15

cgt cat aaa atc gac aag atg tat gtc gcc ttt act gac caa cca att 97
Arg His Lys Ile Asp Lys Met Tyr Val Ala Phe Thr Asp Gln Pro Ile
20 25 30

gag aag gtg caa cag tac act gaa agg gat cgt ttt ttg tct gtc tca 145
Glu Lys Val Gln Gln Tyr Thr Glu Arg Asp Arg Phe Leu Ser Val Ser
35 40 45

gag gcc atg gag ttt ggt ctc ata gat ggg gtg cta gaa aca gaa tac 193

Glu Ala Met Glu Phe Gly Leu Ile Asp Gly Val Leu Glu Thr Glu Tyr
50 55 60

tagttgcaaa tgaatcttta gtagtacatg gtagctagcc ttccaatgac gaaaaagctg 253

gtgttgctca ttaaccactt cgaagtacaa gaagctggct cttgcaaatt tgtatcgtag 313

aaatatctca actcttcaat ccaggaatgt ccaaaagcct aattctgaag acggttatag 373

aaagcgctct tgttttacta tttttgtctc tcctgcagat acactcagca cttttgtggg 433

tattaatcag ggtcttaatt catcacttaa tcacaatcca gttggaagcg aagtgatcaa 493

acacaaagca gattcaggaa gatgtgtatt tttcccaaat atatattact ccaattgcta 553

tcatcccttc gctgtcgtta tgaaaggata tttattttat agcggccgc 602

<210> 22

<211> 64

<212> PRT

<213> Nicotiana tabacum

<400> 22

Arg Pro Leu Glu Asp Val Arg Arg Gln Val Asn Glu Ala Val Gln Pro
1 5 10 15

Arg His Lys Ile Asp Lys Met Tyr Val Ala Phe Thr Asp Gln Pro Ile
20 25 30

Glu Lys Val Gln Gln Tyr Thr Glu Arg Asp Arg Phe Leu Ser Val Ser
35 40 45

Glu Ala Met Glu Phe Gly Leu Ile Asp Gly Val Leu Glu Thr Glu Tyr
50 55 60

<210> 23

<211> 16

<212> DNA

<213> artificial sequence

<220>

<223> primer

<400> 23
agaattcgcg gccgct 16

<210> 24

<211> 32

<212> DNA

<213> artificial sequence

<220>

<223> primer

<400> 24
ctcatgcggc cgcgcgcaac gcaattaatg tg 32

<210> 25

<211> 32

<212> DNA

<213> artificial sequence

<220>

<223> primer

<400> 25
tcatgcggcc gcgagatcca gttcgatgta ac 32

<210> 26

<211> 21

<212> DNA

<213> artificial sequence

<220>

<223> primer

<400> 26
gtggattgat gtgatatctc c 21

<210> 27

<211> 21

<212> DNA

<213> artificial sequence

<220>

<223> primer

<400> 27
gtaaggatct gagctacaca t 21

<210> 28

<211> 24

<212> DNA

<213> artificial sequence

<220>

<223> primer

<400> 28
tataccatgg atttgccatc tttg 24

<210> 29

<211> 21

<212> DNA

<213> artificial sequence

<220>

<223> primer

<400> 29
atagatctca cctggagcca g 21

<210> 30

<211> 19

<212> DNA

<213> artificial sequence

<220>

<223> primer

<400> 30
gagcccatgg caagaggag 19

<210> 31

<211> 22

<212> DNA

<213> artificial sequence

<220>

<223> primer

<400> 31
atagatcttt ctagcttgaa cc 22

<210> 32

<211> 21

<212> DNA

<213> artificial sequence

<220>

<223> primer

<400> 32
tcagccatgg cccctggagg a 21

<210> 33

<211> 24

<212> DNA

<213> artificial sequence

<220>

<223> primer

<400> 33
taagatcttc agtattctgt ttcc 24